

Time-frequency analysis of locally stationary Hawkes processes

F. Roueff¹ and R. von Sachs²

¹*LTCI, Télécom ParisTech, Université Paris-Saclay*

²*Institut de statistique, biostatistique et sciences actuarielles (ISBA) IMMAQ, Université catholique de Louvain*

April 5, 2017

Abstract

Locally stationary Hawkes processes have been introduced in order to generalise classical Hawkes processes away from stationarity by allowing for a time-varying second-order structure. This class of self-exciting point processes has recently attracted a lot of interest in applications in the life sciences (seismology, genomics, neuro-science,...), but also in the modelling of high-frequency financial data. In this contribution we provide a fully developed nonparametric estimation theory of both local mean density and local Bartlett spectra of a locally stationary Hawkes process. In particular we apply our kernel estimation of the spectrum localised both in time and frequency to two data sets of transaction times revealing pertinent features in the data that had not been made visible by classical non-localised approaches based on models with constant fertility functions over time.

Keywords: Time frequency analysis; Locally stationary time series; high frequency financial data; Non-parametric kernel estimation; Self-exciting point processes.

1 Introduction

Many recent time series data modelling and analysis problems increasingly face the challenge of occurring time-variations of the underlying probabilistic structure (mean, variance-covariance, spectral structure,...). This is due to the availability of larger and larger data stretches which can hardly any longer be described by stationary models. Mathematical statisticians ([Dahlhaus \[1996b\]](#), [Zhou and Wu \[2009\]](#), [Birr et al. \[2014\]](#), among others) have contributed with time-localised estimation approaches for many of these time series data based on rigorous models of *locally stationary approximations* to the non-stationary data. Often it has been only via these theoretical studies that well-motivated and fully understood *time-dependent estimation* methods could be developed which correctly adapt to the degree of deviation of the underlying data from a stationary situation (e.g. via the modelling of either a slow - or in contrast a rather abrupt - change of the probabilistic structure over time). For many of these situations, the development of an asymptotic theory of doubly-indexed stochastic processes has proven to be useful: the underlying data stretch is considered to be part of a family of processes which asymptotically approaches a limiting process which locally shows all the characteristics of a stationary process (hence “the locally stationary approximation”). Accompanying estimators ought to adapt to this behavior, e.g. by introduction of local bandwidths in time. Many of these aforementioned approaches have been achieved in the context of classical real-valued series in discrete time, e.g. for (linear) time series via time-varying

MA(∞) representations which apply for a large class of models (see, e.g., [Dahlhaus \[2000\]](#)). Once those are available, the development of a rigorous asymptotic estimation theory is achievable.

This approach, however, is not directly applicable to the model of *locally stationary Hawkes processes*, a class of self-exciting point processes introduced in [Roueff et al. \[2016\]](#) with not only time-varying baseline intensity (such as in [Chen and Hall \[2013\]](#)) but also time-varying fertility function. The fertility function describes the iterative probabilistic mechanism for Hawkes processes to generate offsprings from each occurrence governed by a conditional Poisson point process, given all previous generations. Often, a fertility function $p(t)$ with exponential decay over time t is assumed. The condition $\int p < 1$ ensures a non-explosive accumulation of consecutive populations of the underlying process. When this fertility function is made varying over time through a second argument, one can still rely on the point process mechanism to derive locally stationary approximations.

In this paper we treat estimation of the first and second order structure of locally stationary Hawkes processes on the real line, with a (time-dependent) fertility function $p(\cdot; t)$ assumed to be causal, i.e. supported on \mathbb{R}_+ . Whereas for estimating the local mean density of the process it is sufficient to introduce a localisation via a short window (kernel) in time, for estimating the second-order structure, i.e. the local Bartlett spectrum of the process, one needs to localise both in time and frequency: this time-frequency analysis will be provided via a pair of kernels which concentrate around a given point (t, ω) in the time-frequency plane. Using an appropriate choice of bandwidths in time and in frequency which tend to zero with rates calibrated to minimize the asymptotic mean-square error between the time-frequency estimator and the true underlying local Bartlett spectrum, one can show consistency of our estimator of the latter one. Note that the use of kernels for non-parametric estimation with counting process is not new. To the best of our knowledge this was introduced first by [Ramlau-Hansen \[1983\]](#) for estimating regression point process models such as those introduced in [Aalen \[1975\]](#), see also [Andersen et al. \[1985\]](#) for a general account on the estimation of such processes.

Self exciting point processes have been recently used for modelling point processes resulting from high frequency financial data such as price jump instants (see e.g. [Bacry et al. \[2013\]](#)) or limit order book events (see e.g. [Zheng et al. \[2014\]](#)). In this paper we will illustrate our time frequency analysis approach for point processes on transaction times of two assets (which are Essilor International SA and Total SA).

The rest of the paper is organised as follows. In [Section 2](#), we introduce all necessary preparatory material to develop our estimation theory, including the definition of the population quantities, i.e. local mean density and local Bartlett spectra that we wish to estimate. [Section 3](#) defines our kernel estimators and treats asymptotic bias and variance developments of those under regularity assumptions to be given beforehand. Although these developments are far from being direct and straightforward, the resulting rates of convergence are completely intuitive from a usual nonparametric time-frequency estimation point of view. [Section 4](#) presents the analysis of our transaction data sets leading to interesting observations that have not previously been revealed by a classical analysis with time-constant fertility functions. In [Section 5](#), we present, as a result of independent interest in its own, the necessary new techniques of directly controlling moments of non-stationary Hawkes processes. All proofs are deferred to the Appendix.

2 Preparatory material

In this preparatory section we prepare the ground for developing the presentation of our asymptotic estimation theory of local Bartlett spectra. We do this by introducing a list of useful conventions and definitions, as well as recalling the main concepts and results of [Roueff et al. \[2016\]](#) in as much as they are necessary.

2.1 Conventions and notation

We here set some general conventions and notation adopted all along the paper. Additional ones are introduced in relation with the main assumptions in Section 3.1.

A point process is identified with a random measure with discrete support, $N = \sum_k \delta_{t_k}$ typically, where δ_t is the Dirac measure at point t and $\{t_k\}$ the corresponding (locally finite) random set of points. We use the notation $\mu(g)$ for a measure μ and a function g to express $\int g d\mu$ when convenient. In particular, for a measurable set A , $\mu(A) = \mu(\mathbb{1}_A)$ and for a point process N , $N(g) = \sum_k g(t_k)$. The shift operator of lag t is denoted by S^t . For a set A , $S^t(A) = \{x - t, x \in A\}$ and for a function g , $S^t(g) = g(\cdot + t)$, so that $S^t(\mathbb{1}_A) = \mathbb{1}_{S^t(A)}$. One can then compose a measure μ with S^t , yielding for a function g , $\mu \circ S^t(g) = \mu(g(\cdot + t))$.

We also need some notation for the functional norms which we deal with in this work. Usual L^p -norms are denoted by $|h|_p$,

$$|h|_p = \left(\int |h|^p \right)^{1/p},$$

for $p \in [1, \infty)$ and $|h|_\infty$ is the essential supremum on \mathbb{R} , $|h|_\infty = \text{ess sup}_{t \in \mathbb{R}} |h(t)|$. We also use the following weighted L^p norms which we define to be for any $p \geq 1$, $\beta > 0$, $a \geq 0$ and $h : \mathbb{R} \rightarrow \mathbb{C}$,

$$|h|_{(\beta),p} := |h \times |\cdot|^\beta|_p = \left(\int |h(t)|^\beta dt \right)^{1/p}, \quad (1)$$

$$|h|_{a,p} := |h \times e^{a|\cdot|}|_p = \left(\int |h(t)|^p e^{a p |t|} dt \right)^{1/p}, \quad (2)$$

with the above usual essential sup extensions to the case $p = \infty$.

We denote the convolution product by $*$, that is, for any two functions h_1 and h_2 ,

$$h_1 * h_2(s) = \int h_1(s - t) h_2(t) dt.$$

Finally we use for a random variable X the notation

$$\|X\|_p := (\mathbb{E}|X|^p)^{1/p}, \quad \text{for } p \geq 1. \quad (3)$$

2.2 From stationary to non-stationary and locally stationary Hawkes processes

To start with, we first recall the definition of a stationary (linear) Hawkes process with immigrant intensity λ_c and fertility function p defined on the positive half-line. The conditional intensity function $\lambda(t)$ of such a process is driven by the fertility function taken at the time distances to previous points of the process, i.e. $\lambda(t)$ is given by

$$\lambda(t) = \lambda_c + \int_{-\infty}^{t-} p(t - s) N(ds) = \lambda_c + \sum_{t_i < t} p(t - t_i). \quad (4)$$

Here the first integral is to be read as the integral of the “fertility” function p with respect to the counting process N , which is a sum of Dirac masses at (random) points $(t_i)_{i \in \mathbb{Z}}$. The existence of a stationary point processes with conditional intensity (4) holds under the condition $\int p < 1$ and can be constructed as a cluster point process (see [Daley and Vere-Jones, 2003, EXAMPLE 6.3(c)]).

We extend the stationary Hawkes model defined by the conditional intensity (4) to the non-stationary case by authorizing the immigrant intensity λ_c to be a function $\lambda_c(t)$ of time t and also the fertility function p to be time varying, replacing $p(t-s)$ by the more general $p(t-s; t)$. To ensure a locally finite point process in this definition, we impose the two conditions

$$\zeta_1 := \sup_{t \in \mathbb{R}} \int p(s; t) ds < 1 \quad \text{and} \quad |\lambda_c|_\infty < \infty. \quad (5)$$

They yield the existence of a non-stationary point process N with a mean density function which is uniformly bounded by $|\lambda_c|_\infty / (1 - \zeta_1)$ (see [Roueff et al., 2016, Definition 1]).

As non-stationary Hawkes processes can evolve quite arbitrarily over time, the statistical analysis of them requires to introduce local stationary approximations in the same fashion as time varying autoregressive processes in time series, for which *locally stationary* models have been successfully introduced (see Dahlhaus [1996b]). Thus, a locally stationary Hawkes process with *local immigrant intensity* $\lambda_c^{<LS>}(\cdot)$ and *local fertility function* $p^{<LS>}(\cdot; \cdot)$ is a collection $(N_T)_{T \geq 1}$ of non-stationary Hawkes processes with respective immigrant intensity and fertility function given by $\lambda_{cT}(t) = \lambda_c^{<LS>}(t/T)$ and varying fertility function given by $p_T(\cdot; t) = p^{<LS>}(\cdot; t/T)$, see [Roueff et al., 2016, Definition 2] where this model is called a *locally stationary* Hawkes process. For a given *real location* t , the scaled location t/T is typically called an *absolute location* and denoted by u or v .

Note that the collection $(N_T)_{T \geq 1}$ of non-stationary Hawkes processes are defined using the same time varying parameters $\lambda_c^{<LS>}$ and $p^{<LS>}$ but with the time varying arguments scaled by T . As a result, the larger T is, the slower the parameters evolve along the time.

An assumption corresponding to (5) to guarantee that, for all $T \geq 1$, the non-stationary Hawkes process N_T admits a uniformly bounded mean density function is the following:

$$\zeta_1^{<LS>} := \sup_{u \in \mathbb{R}} \int p^{<LS>}(r; u) dr < 1 \quad \text{and} \quad |\lambda_c^{<LS>}|_\infty < \infty. \quad (6)$$

Under this assumption, for each absolute location $u \in \mathbb{R}$, the function $r \mapsto p^{<LS>}(r; u)$ satisfies the required condition for the fertility function of a stationary Hawkes process. Hence, assuming (6), for any absolute location u , we denote by $N(\cdot; u)$ a stationary Hawkes process with constant immigrant intensity $\lambda_c^{<LS>}(u)$ and fertility function $r \mapsto p^{<LS>}(r; u)$. In the following subsection we will include this assumption (6) into a stronger set of assumptions that we use for derivation of the results on asymptotic estimation theory.

We also remark that, for any $T \geq 1$, the conditional intensity function λ_T of the non-stationary Hawkes process N_T takes the form

$$\begin{aligned} \lambda_T(t) &= \lambda_c^{<LS>}(t/T) + \int_{-\infty}^{t^-} p^{<LS>}(t-s; t/T) N_T(ds) \\ &= \lambda_c^{<LS>}(t/T) + \sum_{t_{i,T} < t} p^{<LS>}(t-t_{i,T}; t/T), \end{aligned}$$

where $(t_{i,T})_{i \in \mathbb{Z}}$ denote the points of N_T . This latter formula can also be used to simulate locally stationary Hawkes processes on the real line. The examples for locally stationary Hawkes processes (with time varying Gamma shaped fertility functions) used in [Roueff et al., 2016, Section 2.6] were simulated in this way.

First and second order statistics for point processes are of primary importance for statistical inference. As for time series they are conveniently described in the stationary case by a mean parameter for the first order statistics and a spectral representation, the so called Bartlett spectrum (see [Daley and Vere-Jones, 2003, Proposition 8.2.I]), for the covariance structure. The locally stationary approach allows us to define such quantities as depending on the absolute time u as introduced in the following section.

2.3 Local mean density and Bartlett spectrum

Consider a locally stationary Hawkes process $(N_T)_{T \geq 1}$ with local immigrant intensity $\lambda_c^{<LS>}$ and local fertility function $p^{<LS>}(\cdot; \cdot)$ satisfying condition (6). Although for a given T , the first and second order statistics of N_T can be quite involved, some intuitive asymptotic approximations are available as T grows to infinity. Namely, for any absolute time u , the *local* statistical behavior of N_T around real time Tu has to be well approximated by that of a stationary Hawkes process with (constant) immigrant intensity $\lambda_c^{<LS>}(u)$ and fertility function $p^{<LS>}(\cdot; u)$. This stationary Hawkes process at absolute location u is denoted in the following by $N(\cdot; u)$. Precise approximation results are provided in Roueff et al. [2016] and recalled in Section 3.3.1 below. Presently, we only need to introduce how to define this local first and second order statistical structure.

We first introduce the *local mean density* function $m_1^{<LS>}(u)$ defined at each absolute location u , as the mean parameter of the stationary Hawkes process $N(\cdot; u)$. By [Daley and Vere-Jones, 2003, Eq. (6.3.26) in Example 6.3(c)], it is given by

$$m_1^{<LS>}(u) = \frac{\lambda_c^{<LS>}(u)}{1 - \int p^{<LS>}(\cdot; u)} . \quad (7)$$

A convenient way to describe the covariance structure of a second order stationary point process N on \mathbb{R} is to rely on a spectral representation, the Bartlett spectrum, which is defined as the (unique) non-negative measure Γ on the Borel sets such that, for any bounded and compactly supported function f on \mathbb{R} , (see [Daley and Vere-Jones, 2003, Proposition 8.2.I])

$$\text{Var}(N(f)) = \Gamma(|F|^2) = \int |F(\omega)|^2 \Gamma(d\omega) ,$$

where F denotes the Fourier transform of f ,

$$F(\omega) = \int f(t) e^{-it\omega} dt .$$

For the stationary Hawkes processes $N(\cdot; u)$, the Bartlett spectrum admits a density given by (see [Daley and Vere-Jones, 2003, Example 8.2(e)])

$$\gamma^{<LS>}(u; \omega) = \frac{m_1^{<LS>}(u)}{2\pi} \left| 1 - P^{<LS>}(\omega; u) \right|^{-2} , \quad (8)$$

where

$$P^{<LS>}(\omega; u) = \int p^{<LS>}(t; u) e^{-it\omega} dt .$$

Analagous to the first order structure, we call $\gamma^{<LS>}(u; \omega)$ the *local Bartlett spectrum density* at frequency ω and absolute location u . This local Bartlett spectrum density plays a role similar to that of the local spectral density $f(u, \lambda)$ introduced in [Dahlhaus, 1996b, Page 142] for locally stationary time series.

2.4 Estimators

As our approach is local in time and frequency, we rely on two kernels w and q which are required to be compactly supported (see Remark 1 below). More precisely, we have the following assumptions.

- (K-1) Let w be a $\mathbb{R} \rightarrow \mathbb{R}_+$ bounded function with compact support such that $\int w = |w|_1 = 1$.
- (K-2) Let q be a $\mathbb{R} \rightarrow \mathbb{C}$ bounded function with compact support such that $|q|_2 = 1$.

To localize in time let $b_1 > 0$ be a given time bandwidth and define w_{b_1} and w_{Tb_1} to be the corresponding kernels in absolute time u and real time t , namely,

$$w_{b_1}(u) := b_1^{-1} w(u/b_1) \quad \text{and} \quad w_{Tb_1}(t) := T^{-1} w_{b_1}(t/T) = (Tb_1)^{-1} w(t/(Tb_1)) . \quad (9)$$

Let now u_0 be a fixed absolute time. For estimating the local mean density $m_1^{<LS>}(u_0)$ given by equation (7), approximating the mean density function $t \mapsto m_{1T}(t)$ of N_T locally in the neighborhood of Tu_0 by $m_1^{<LS>}(u_0)$ we have, for Tb_1 small,

$$m_1^{<LS>}(u_0) \approx \int w_{Tb_1}(t - Tu_0) m_{1T}(t) dt ,$$

where we used $\int w_{Tb_1} = 1$ and that the support of $t \mapsto w_{Tb_1}(t - Tu_0)$ essentially lives in the neighborhood of Tu_0 for Tb_1 small. Since the right-hand side of this approximation is $\mathbb{E} [N_T(S^{-Tu_0} w_{Tb_1})]$, this suggests the following estimator of $m_1^{<LS>}(u_0)$,

$$\hat{m}_{b_1}(u_0) := N_T(S^{-Tu_0} w_{Tb_1}) = \int w_{Tb_1}(t - Tu_0) N_T(dt) . \quad (10)$$

For estimation of the second order structure, i.e. the local Bartlett spectral density $\gamma^{<LS>}(u_0; \omega_0)$ for some given point (u_0, ω_0) of the time-frequency plane, we need also to localise in frequency by a kernel which will be given by the (squared) Fourier transform $|Q|^2$ of the kernel q . Then for a given frequency bandwidth $b_2 > 0$, we are looking for an estimator of the auxiliary quantity

$$\gamma_{b_2}^{<LS>}(u_0; \omega_0) := \int \frac{1}{b_2} \left| Q\left(\frac{\omega - \omega_0}{b_2}\right) \right|^2 \gamma^{<LS>}(\omega; u_0) d\omega , \quad (11)$$

which in turn, as $b_2 \rightarrow 0$, is an approximation of the density $\gamma^{<LS>}(u_0; \omega_0)$. The rate of approximation (i.e. the ‘‘bias in frequency direction’’ of the following estimator) is established in Theorem 2, equation (27), below. Let us now set $q_{\omega_0, b_2}(t) = b_2^{1/2} e^{i\omega_0 t} q(b_2 t)$ such that the squared modulus of its Fourier transform writes as

$$|Q_{\omega_0, b_2}(\omega)|^2 = \frac{1}{b_2} \left| Q\left(\frac{\omega - \omega_0}{b_2}\right) \right|^2 .$$

Using that $\gamma^{<LS>}(\omega; u_0) d\omega$ is the Bartlett spectrum of $N(\cdot; u_0)$ as recalled in Section 2.3, we can thus rewrite (11) as

$$\gamma_{b_2}^{<LS>}(u_0; \omega_0) = \text{Var} (N(q_{\omega_0, b_2}; u_0)) . \quad (12)$$

Since this variance is an approximation of $\text{Var} (N_T(S^{-Tu_0} q_{\omega_0, b_2}))$, where q_{ω_0, b_2} is shifted to be localized around Tu_0 , we finally estimate $\gamma_{b_2}^{<LS>}(u_0; \omega_0)$ by the following moment estimator:

$$\hat{\gamma}_{b_2, b_1}(u_0; \omega_0) := \hat{E} [|N_T(S^{-Tu_0} q_{\omega_0, b_2})|^2; w_{b_1}] - \left| \hat{E} [N_T(S^{-Tu_0} q_{\omega_0, b_2}); w_{b_1}] \right|^2 , \quad (13)$$

where for a test function f localized around the origin and taking $\rho(x) = x$ and $\rho(x) = |x|^2$ successively, we have built an estimator of $\mathbb{E}[\rho(N(f; u_0))]$ based on the empirical observations of N_T and defined by

$$\begin{aligned} \hat{E}[\rho(N_T(S^{-Tu_0} f)); w_{b_1}] &:= \int \rho(N_T(f(\cdot - t - Tu_0))) w_{Tb_1}(t) dt \\ &= \frac{1}{T} \int \rho\left(\sum_k f(t_{k,T} - t - Tu_0)\right) w_{b_1}(t/T) dt . \end{aligned} \quad (14)$$

Note that in (14) the dependence of the estimator on u_0 appears in the composition of the observed point process N_T with the shift operator S^{-Tu_0} . By an obvious change of variable, this would be equivalent to let N_T unshifted as $N_T(f(\cdot - t))$ and instead shift $w_{Tb_1}(t)$ into $w_{Tb_1}(t - Tu_0)$, or, in absolute time, shift $w_{b_1}(u)$ into $w_{b_1}(u - u_0)$.

Remark 1. In practice N_T is observed over a finite interval. In order to have estimators $\hat{m}_{b_1}(u_0)$ and $\hat{\gamma}_{b_2, b_1}(u_0; \omega_0)$ in (10) and (13) that only use observations within this interval, the supports of w and q must be bounded and some restriction imposed on b_1 , b_2 and T . Suppose for instance that N_T is observed on $[0, T]$ (mimicking the usual convention for locally stationary time series of [Dahlhaus \[1996a\]](#)). The local mean density and Bartlett spectrum can then be estimated at corresponding absolute times $u_0 \in (0, 1)$ and the restrictions on b_1 , b_2 and T read as follows. In (10), we must have $u_0 + b_1 \text{Supp}(w) \subseteq [0, 1]$, and in (13), we must have $u_0 + b_1 \text{Supp}(w) + (Tb_2)^{-1} \text{Supp}(q) \subseteq [0, 1]$. These two support conditions are always satisfied, eventually as $T \rightarrow \infty$, provided that the kernels w and q are compactly supported and that $b_1 \rightarrow 0$ and $Tb_2 \rightarrow \infty$.

In the sequel we will show that this is a sensible estimator of $\gamma_{b_2}^{<LS>}(u_0; \omega_0)$ sharing the usual properties of a nonparametric estimator constructed via kernel-smoothing over time and frequency: for sufficiently small bandwidths b_1 in time and b_2 in frequency this estimator becomes well localised around $(u_0; \omega_0)$.

The main results stated hereafter provide asymptotic expansions of its bias and variance behaviour, leading to consistency of this estimator under some asymptotic condition for b_1 and b_2 as $T \rightarrow \infty$.

3 Bias and variance bounds

3.1 Main assumptions

The first assumption is akin but stronger than condition (6) above, being in fact equal to assumption (LS-1) of [Roueff et al. \[2016\]](#). It guarantees that, for all $T \geq 1$, the locally stationary Hawkes process N_T admits a (causal) local fertility function $s \mapsto p^{<LS>}(s; u)$ which is not only uniformly bounded, but has an exponentially decaying memory (as a function in the first argument, uniformly with respect to its second argument).

(LS-1) Assume that

$$|\lambda_c^{<LS>}|_\infty < \infty. \quad (15)$$

Assume moreover that for all $u \in \mathbb{R}$, $p^{<LS>}(\cdot; u)$ is supported on \mathbb{R}_+ and that there exists a $d > 0$ such that $\zeta_1^{<LS>}(d) < 1$ and $\zeta_\infty^{<LS>}(d) < \infty$ where

$$\zeta_1^{<LS>}(d) := \sup_{u \in \mathbb{R}} |p^{<LS>}(\cdot; u)|_{d,1} = \sup_{u \in \mathbb{R}} \int p^{<LS>}(s; u) e^{d|s|} ds \quad (16)$$

and

$$\zeta_\infty^{<LS>}(d) := \sup_{u \in \mathbb{R}} |p^{<LS>}(\cdot; u)|_{d,\infty} = \sup_{u \in \mathbb{R}} \text{ess sup}_{s \in \mathbb{R}} \left\{ |p^{<LS>}(s; u)| e^{d|s|} \right\}. \quad (17)$$

All the examples considered in [Roueff et al. \[2016\]](#) satisfy this condition. It is also satisfied if the local fertility functions have a (uniformly) bounded compact support (cf. [Hansen et al. \[2015\]](#) in the stationary case).

(LS-2) Assume that, for some $\beta \in (0, 1]$,

$$\xi_c^{(\beta)} := \sup_{u \neq v} \frac{|\lambda_c^{<LS>}(v) - \lambda_c^{<LS>}(u)|}{|v - u|^\beta} < \infty.$$

(LS-3) Assume that, for some $\beta \in (0, 1]$, $|\xi^{(\beta)}|_1 < \infty$, where

$$\xi^{(\beta)}(r) := \sup_{u \neq v} \frac{|p^{<LS>}(r; v) - p^{<LS>}(r; u)|}{|v - u|^\beta}.$$

These assumptions can be interpreted as smoothness conditions on $\lambda_c^{<LS>}$ ((LS-2)) and on $p^{<LS>}(\cdot; \cdot)$ with respect to its second argument ((LS-3)). Note also that Assumptions (16) and (17) imply in particular Assumption (LS-4) of Roueff et al. [2016] which we recall here to be

$$\zeta_\infty^{<LS>} := \zeta_\infty^{<LS>}(0) = \sup_{u \in \mathbb{R}} |p^{<LS>}(\cdot; u)|_\infty < \infty, \quad (18)$$

$$\zeta_{(\beta)}^{<LS>} := \sup_{u \in \mathbb{R}} |p^{<LS>}(\cdot; u)|_{(\beta),1} < \infty. \quad (19)$$

This can be seen simply by noting that $\zeta_1^{<LS>} := \zeta_1^{<LS>}(0) \leq \zeta_1^{<LS>}(d)$ and $\zeta_\infty^{<LS>} \leq \zeta_\infty^{<LS>}(d)$ for all $d \geq 0$, with equality for $d = 0$. Similarly, $\zeta_1^{<LS>}(d) < \infty$ for some $d > 0$ implies $\zeta_{(\beta)}^{<LS>} < \infty$ for all $\beta > 0$.

Hereafter all the given bounds are minimax upper bounds in the sense that they hold uniformly over parameters $\lambda_c^{<LS>}$ and $p^{<LS>}$ satisfying the set of conditions (6), (LS-2), (LS-3), and (18) and (19), as in Theorems 1, or the more restrictive set of conditions (LS-1), (LS-2) and (LS-3), as in Theorem 2 and 3. More specifically, we use the following conventions all along the paper.

Convention 1 (Symbol \lesssim). For two nonnegative sequences a_T and b_T indexed by $T \geq 1$, possibly depending on parameters $\lambda_c^{<LS>}$, $p^{<LS>}$, b_1 and b_2 , we use the notation $a_T \lesssim b_T$ to denote that there exists a constant C such that, for all b_1, b_2 and T satisfying certain conditions $\mathcal{C}(b_1, b_2, T)$, we have $a_T \leq C b_T$ with C only depending on non-asymptotic quantities and constants such as d , β , $\zeta_1^{<LS>}(d)$, $\zeta_{(\beta)}^{<LS>}$, $\zeta_\infty^{<LS>}(d)$, $\xi_c^{(\beta)}$, $|\xi^{(\beta)}|_1$, $|\lambda_c^{<LS>}|_\infty$ and the two kernel functions w and q .

The conditions $\mathcal{C}(b_1, b_2, T)$ will be intersections of the following ones :

$$T \geq 1 \quad \text{and} \quad b_1 \in (0, 1], \quad (20)$$

$$b_1, b_2 \in (0, 1] \quad \text{and} \quad T b_1 b_2 \geq 1, \quad (21)$$

$$b_1 \ln(T) \leq 1. \quad (22)$$

Convention 2 (Constants A_1, A_2). We use A_1, A_2 to denote positive constants that can change from one expression to another but always satisfy $A_1^{-1} \lesssim 1$ and $A_2 \lesssim 1$, using Convention 1. In other words A_1 and A_2 are positive constants which can be bounded from below and from above, respectively, using the constants appearing in the assumptions and the chosen kernels w and q .

Convention 2 will be useful to treat exponential terms in a simplified way, that is, without considering unnecessary constants; for instance, we can write $(e^{-A_1 T})^2 \leq e^{-A_1 T}$ replacing $2A_1$ by A_1 in the second expression without affecting the property $A_1^{-1} \lesssim 1$ required on A_1 .

3.2 Main results

We can now state the main results of this contribution, whose proofs can be found in Appendix C. For the bias and variance of the local mean density estimator $\hat{m}_{b_1}(u_0)$ we establish the following result.

Theorem 1. *Let the kernel w satisfy (K-1). Assume conditions (6), (LS-2), (LS-3), and (18) and (19) to hold. Then, for b_1 and T satisfying (20), the bias of the local density estimator satisfies, for all $u_0 \in \mathbb{R}$,*

$$|\mathbb{E}[\hat{m}_{b_1}(u_0)] - m_1^{<LS>}(u_0)| \lesssim b_1^\beta + T^{-\beta}. \quad (23)$$

If moreover (LS-1) holds, its variance satisfies

$$\text{Var}(\hat{m}_{b_1}(u_0)) \lesssim (Tb_1)^{-1}. \quad (24)$$

Hence, $\hat{m}_{b_1}(u_0)$ is shown to be a (mean-square) consistent estimator of $m_1^{<LS>}(u_0)$, and, optimizing the bias and variance bounds, we get the “usual” mean-square error rate $T^{-\frac{2\beta}{2\beta+1}}$ for nonparametric curve estimation with an additive noise structure, achieved for a bandwidth $b_1 \sim T^{-\frac{1}{2\beta+1}}$.

We now treat the bias of the estimator $\hat{\gamma}_{b_2, b_1}(u_0; \omega_0)$ which can be decomposed as the sum of 1) a bias in the time direction, namely, $\mathbb{E} \hat{\gamma}_{b_2, b_1}(u_0; \omega_0) - \gamma_{b_2}^{<LS>}(u_0; \omega_0)$ and 2) a bias in the frequency direction, namely, $\gamma_{b_2}^{<LS>}(u_0; \omega_0) - \gamma^{<LS>}(u_0; \omega_0)$.

Theorem 2. *Let the kernels w and q satisfy (K-1) and (K-2). Assume conditions (LS-1), (LS-2), and (LS-3) to hold. Then, for all b_1, b_2 and T satisfying (21) and (22) and for all $u_0, \omega_0 \in \mathbb{R}$, we have*

$$|\mathbb{E}[\hat{\gamma}_{b_2, b_1}(u_0; \omega_0)] - \gamma_{b_2}^{<LS>}(u_0; \omega_0)| \lesssim b_1^\beta + b_1^{2\beta} b_2^{-1} + (Tb_1 b_2)^{-1}. \quad (25)$$

If moreover the squared modulus $|Q(\omega)|^2$ of the Fourier transform of the kernel q satisfies

$$\int \omega^2 |Q(\omega)|^2 d\omega < \infty \quad \text{and} \quad \int \omega |Q(\omega)|^2 d\omega = 0, \quad (26)$$

the “bias in frequency direction” fulfills for $b_2 \in (0, 1]$,

$$|\gamma_{b_2}^{<LS>}(u_0; \omega_0) - \gamma^{<LS>}(u_0; \omega_0)| \lesssim b_2^2. \quad (27)$$

Remark 2. *Condition (26) is automatically satisfied if q is compactly supported, real valued and even, and admits an L^2 derivative, such as the triangle shape kernel.*

We already observe here that for the estimator $\hat{\gamma}_{b_2, b_1}(u_0; \omega_0)$ to be asymptotically unbiased, equations (25) and (27) require the following conditions on the choice of the two bandwidths b_1 and b_2 to be fulfilled:

$$Tb_1 b_2 \rightarrow \infty, \quad b_2 \rightarrow 0 \quad \text{and} \quad b_2^{-1} b_1^{2\beta} \rightarrow 0.$$

Note in particular that these conditions for an asymptotically unbiased estimator imply those required for the feasibility of the estimator in Remark 1 ($b_1 \rightarrow 0$ and $Tb_2 \rightarrow \infty$).

We shall discuss possible compatible bandwidth choices below, following the treatment of the variance of this estimator.

Theorem 3. *Let the kernels w and q satisfy (K-1) and (K-2). Assume conditions (LS-1), (LS-2) and (LS-3) to hold. Then, for all b_1, b_2, T satisfying (21), and for all $u_0, \omega_0 \in \mathbb{R}$, we have*

$$\text{Var}(\hat{\gamma}_{b_2, b_1}(u_0; \omega_0)) \lesssim (Tb_1 b_2)^{-1} + b_1^{2\beta} (b_1^\beta b_2^{-1})^2. \quad (28)$$

In order to optimize bandwidth choices in time and in frequency to derive an optimal rate of MSE-consistency of the estimator $\hat{\gamma}_{b_2, b_1}(u_0; \omega_0)$ for a given time-frequency point $(u_0; \omega_0)$ we observe, by equations (25), (27) and (28), that the MSE satisfies

$$\text{MSE}(\hat{\gamma}_{b_2, b_1}(u_0; \omega_0)) \lesssim b_2^4 + b_1^{2\beta} + (Tb_1 b_2)^{-1} + (b_1^{2\beta} b_2^{-1})^2.$$

The MSE-rate $T^{-\frac{4\beta}{5\beta+2}}$ is achieved by optimizing this upper bound, that is, by imposing $b_2^4 \sim b_1^{2\beta} \sim (Tb_1 b_2)^{-1}$ (leading to $(b_1^{2\beta} b_2^{-1})^2 \sim b_2^6$), and hence by setting $b_1 \sim T^{-\frac{2}{2+5\beta}}$ and $b_2 \sim T^{-\frac{\beta}{2+5\beta}}$.

3.3 Main ideas of the proofs

3.3.1 Local approximations of moments

An essential step for treating the bias terms is to be able to approximate, as $T \rightarrow \infty$, in the neighborhood of uT , the first and second moments of N_T by that of the local stationary approximation $N(\cdot; u)$ defined as in Section 2.3. We first state the two approximations that directly follow [Roueff et al., 2016, Theorem 4] with $m = 1, 2$.

Theorem 4. *Let $\beta \in (0, 1]$ and let $(N_T)_{T \geq 1}$ be a locally stationary Hawkes process satisfying conditions (6), (LS-2), (LS-3), (18) and (19). Let $(N(\cdot; u))_{u \in \mathbb{R}}$ be the collection of stationary Hawkes process defined as in Section 2.3. Then, for all $T \geq 1$, $u \in \mathbb{R}$ and all bounded integrable functions g , we have*

$$|\mathbb{E}[N_T(S^{-Tu}g)] - \mathbb{E}[N(g; u)]| \lesssim (|g|_1 + |g|_{(\beta),1}) T^{-\beta}, \quad (29)$$

$$|\text{Var}(N_T(S^{-Tu}g)) - \text{Var}(N(g; u))| \lesssim (|g|_1 + |g|_\infty) (|g|_1 + |g|_{(\beta),1}) T^{-\beta}. \quad (30)$$

The control of the bias in Theorem 1 directly follows from (29). However it turns out that (30) is not sharp enough to control the bias of the local Bartlett spectrum and thus to obtain the expected convergence rate. The basic reason is that it involves L^1 (weighted) norms of g in the upper bound instead of L^2 norms. In order to recover the correct rates of convergence, we rely on the following new result where the L^1 (weighted) norms are indeed replaced by L^2 (weighted) norms, or, to be more precise, where the remaining weighted L^1 -norms are compensated by an exponentially decreasing term in T , and will thus turn out to be negligible.

Theorem 5. *Let $\beta \in (0, 1]$ and let $(N_T)_{T \geq 1}$ be a locally stationary Hawkes process satisfying conditions (LS-1), (LS-2) and (LS-3). Let $(N(\cdot; u))_{u \in \mathbb{R}}$ be the collection of stationary Hawkes process defined as in Section 2.3. Then, for all bounded and compactly supported functions g , for all $T \geq 1$ and $u \in \mathbb{R}$,*

$$|\text{Var}(N_T(g))| \lesssim |g|_2^2 + e^{-A_1 T} |g|_{d,1}^2, \quad (31)$$

$$\begin{aligned} & |\text{Var}(N_T(S^{-Tu}g)) - \text{Var}(N(g; u))| \\ & \lesssim \left\{ |g|_2^2 + e^{-A_1 T} |g|_{d,1}^2 + |g|_{(\beta),2} \left(|g|_2 + e^{-A_1 T} |g|_{d,1} \right) \right\} T^{-\beta}. \end{aligned} \quad (32)$$

The proof of this theorem can be found in Appendix B. To obtain this new result, we crucially rely on the assumption (LS-1) where controls in exponentially weighted norms (based on $\sup_u |p^{<LS>}(\cdot; u)|_{d,q}$ for $q = 1, \infty$ and some $d > 0$) are assumed to strengthen the assumptions (6), (18) and (19).

3.3.2 Bias and approximate centering

To control all the error terms, we found useful to introduce a centered version of N_T with a centering term corresponding to its asymptotic deterministic version. Recalling that N_T behaves in a neighborhood of $T u_0$ as $N(\cdot; u_0)$ and that this process admits the mean intensity denoted by $m_1^{<LS>}(u_0)$, we define, for any test function f ,

$$\overline{N}_T(f; u_0) := N_T(f) - \mathbb{E}[N(f; u_0)] = \int f(s) [N_T(ds) - m_1^{<LS>}(u_0) ds]. \quad (33)$$

It is important to note that this “approximate” centering depends on an absolute location u_0 as it is a good approximation of $\mathbb{E}[N_T(f)]$ only for f localized in a neighborhood of

$T u_0$. Let us apply this definition. By (10), since w_{Tb_1} integrates to 1, the error of the local mean density estimator can directly be expressed as

$$\widehat{m}_{b_1}(u_0) - m_1^{<LS>}(u_0) = \overline{N}_T(S^{-Tu_0} w_{Tb_1}; u_0). \quad (34)$$

Hence controlling the bias of this estimator directly amounts to evaluating the quality of the above centering.

The treatment of the bias of the local Bartlett spectrum is a bit more involved since, as often for spectral estimators, the empirical centering term requires a specific attention. This term appears inside the negated square modulus of the right-hand side of (13). To see why it is indeed a centering term, observe that, using $\int w_{Tb_1} = 1$, we can write

$$\widehat{\gamma}_{b_2, b_1}(u_0; \omega_0) = \widehat{E} [\rho(N_T(S^{-Tu_0} q_{\omega_0, b_2})); w_{b_1}]$$

with now ρ defined as the “centered” squared modulus

$$\rho(x) = \left| x - \widehat{E} [N_T(S^{-Tu_0} q_{\omega_0, b_2}); w_{b_1}] \right|^2.$$

Using that, for the above definition of ρ , we have, for any test function f ,

$$\rho(N_T(f)) = \left| \overline{N}_T(f; u_0) - \widehat{E} [\overline{N}_T(f; u_0); w_{b_1}] \right|^2.$$

By comparison with the previous expression of $\widehat{\gamma}_{b_2, b_1}(u_0; \omega_0)$, we easily get an expression of the local Bartlett spectrum estimator based on this centered version of N_T , namely,

$$\widehat{\gamma}_{b_2, b_1}(u_0; \omega_0) = \widehat{E} [|\overline{N}_T(f; u_0)|^2; w_{b_1}] - \left| \widehat{E} [\overline{N}_T(f; u_0); w_{b_1}] \right|^2, \quad (35)$$

where analogously to (14), we define, for the test function $f = S^{-Tu_0} q_{\omega_0, b_2}$,

$$\begin{aligned} \widehat{E} (|\overline{N}_T(f; u_0)|^2; w_{b_1}) &:= \int |\overline{N}_T(f(\cdot - t); u_0)|^2 w_{Tb_1}(t) dt, \\ \widehat{E} (\overline{N}_T(f; u_0); w_{b_1}) &:= \int \overline{N}_T(f(\cdot - t); u_0) w_{Tb_1}(t) dt. \end{aligned} \quad (36)$$

In fact, using $\int w = 1$, f integrable and interchanging the order of integration, we immediately get the simplification

$$\widehat{E} (\overline{N}_T(f; u_0); w_{b_1}) = \overline{N}_T(f * w_{Tb_1}; u_0), \quad (37)$$

where we used the convolution product $*$. The advantage of the new expression (35) in contrast to the original (13) is that now we expect the negated square modulus to be of negligible order. To see why, consider for instance the bias in (25), for the control of which we need to bound, as the second term of (35),

$$\mathbb{E} \left[\left| \widehat{E} (\overline{N}_T(f; u_0); w_{b_1}) \right|^2 \right] = \text{Var} (N_T(f * w_{Tb_1})) + \left| \mathbb{E} [\overline{N}_T(f * w_{Tb_1}; u_0)] \right|^2. \quad (38)$$

(Here and in the following we repeatedly use the fact that N_T and $\overline{N}_T(\cdot; u_0)$ have the same variance). Finally, the control of the bias term in (25) now requires to evaluate

$$\mathbb{E} \left[\widehat{E} [|\overline{N}_T(S^{-Tu_0} q_{\omega_0, b_2}; u_0)|^2; w_{b_1}] \right] - \gamma_{b_2}^{<LS>}(u_0; \omega_0),$$

which is again decomposed as a main term

$$\int (\text{Var} (N_T(S^{-Tu_0} q_{\omega_0, b_2}(\cdot - t))) - \gamma_{b_2}^{<LS>}(u_0; \omega_0)) w_{Tb_1}(t) dt, \quad (39)$$

added to a negligible (because involving a squared bias) term

$$\int \left| \mathbb{E} [\overline{N}_T(S^{-Tu_0} q_{\omega_0, b_2}(\cdot - t); u_0)] \right|^2 w_{Tb_1}(t) dt. \quad (40)$$

3.3.3 Variance terms and exact centering

The variance of the local mean density estimator directly requires to control the variance of $N_T(f)$ for given test functions f . This requires new deviation bounds for non-stationary Hawkes processes. By deviation bounds we here mean that we bound the moments of

$$\overline{\overline{N}}_T(h) := N_T(h) - \mathbb{E}[N_T(h)] , \quad (41)$$

where h is an appropriate test function. New results in this direction are gathered in Section 5, where the dependence structure of non-stationary Hawkes processes is investigated, leading to the appropriate control of such moments in Proposition 9. For instance the moment of order 2 directly provides the adequate bound for the variance of the local mean density estimator

$$\text{Var}(\widehat{m}_{b_1}(u_0)) = \text{Var}(N_T(S^{-T u_0} w_{T b_1})) = \left\| \overline{\overline{N}}_T(h) \right\|_2^2 , \quad (42)$$

where we use the notation introduced in (3).

Now we turn our attention to the estimator of the local Bartlett spectrum. The control of the moments of $\overline{\overline{N}}_T$ will essentially be used to approximate $\widehat{\gamma}_{b_2, b_1}(u_0; \omega_0)$ by

$$\widetilde{\gamma}_{b_2, b_1}(u_0; \omega_0) := \int \left| \overline{\overline{N}}_T(S^{-T u_0} q_{\omega_0, b_2}(\cdot - t)) \right|^2 w_{T b_1}(t) dt . \quad (43)$$

In contrast to the centering used in $\overline{N}_T(\cdot; u)$ for controlling the bias (a centering with respect to $\mathbb{E}[N(h; u)]$) for some absolute location u , here the term $\mathbb{E}[N_T(h)]$ is no longer invariant as h is shifted. This is why this centering cannot be used as a direct decomposition of estimator $\widehat{\gamma}_{b_2, b_1}(u_0; \omega_0)$ as in (35). Instead we use a bound on the error of approximating $\widehat{\gamma}_{b_2, b_1}(u_0; \omega_0)$ by $\widetilde{\gamma}_{b_2, b_1}(u_0; \omega_0)$, see Lemma 16.

Finally the variance of the local Bartlett estimator is obtained by controlling the variance of $\widetilde{\gamma}_{b_2, b_1}(u_0; \omega_0)$ (Lemma 15), which in turn relies on a bound of

$$\text{Cov} \left(\left| \overline{\overline{N}}_T(h_1) \right|^2 , \left| \overline{\overline{N}}_T(h_2) \right|^2 \right)$$

for test functions h_1 and h_2 , which is derived in Corollary 10.

4 Numerical experiments

The numerical experiments in Roueff et al. [2016] show that the estimators \widehat{m}_{b_1} and $\widehat{\gamma}_{b_2, b_1}$ are able to reproduce the theoretical local mean density and local Bartlett spectrum on simulated locally stationary Hawkes processes. Here we consider a real data set containing the transaction times of the two assets ESSI.PA (Essilor International SA) and TOTF.PA (Total SA) over 61 days scattered in February, June and November 2013. We computed the local mean density and Bartlett spectrum estimators, say $\widehat{m}_{b_1}^{(i)}$ and $\widehat{\gamma}_{b_2, b_1}^{(i)}$ for each day $i \in \{1, \dots, 61\}$ over the regular opening hours of the Paris stock exchange market, that is between 9:00 a.m. and 5:30 p.m., Paris local time. The estimators are computed with the following kernels : w is the triangle kernel and q is the Epanechnikov kernel, both with supports $[-.5, .5]$. The chosen bandwidth parameters are given by

$$b_1 = .15 , \quad b_2 = .005 \text{ Hz} .$$

The above unit for b_1 is absolute time, that is, 1 unit corresponds to the overall duration of observation $T = 8.5$ hours, hence in real time, $b_1 = .15 * 8.5$ hours, which makes 1 hour, 16 minutes and 30 seconds. We thus obtain for each asset 61 local mean density and local

Bartlett spectrum estimates. Our goal here is to illustrate the time frequency analysis of such point processes data sets. The obtained results are quite different from one day to another, which can be expected on such real data. However, in the following, we propose to comment on the local mean densities and Bartlett spectra obtained for the two assets by averaging over the available 61 days,

$$\hat{m}_{b_1}^{(Av)} = \frac{1}{61} \sum_{i=1}^{61} \hat{m}_{b_1}^{(i)} \quad \text{and} \quad \hat{\gamma}_{b_2, b_1}^{(Av)} = \frac{1}{61} \sum_{i=1}^{61} \hat{\gamma}_{b_2, b_1}^{(i)} .$$

Moreover we computed a *Poisson-normalized* local Bartlett spectrum of these averaged estimates defined by

$$\hat{\gamma}_{b_2, b_1}^{(Pn)}(\omega; u) = \frac{2\pi \hat{\gamma}_{b_2, b_1}^{(Av)}(\omega; u)}{\hat{m}_{b_1}^{(Av)}(u)} \quad u \in \mathbb{R}, \omega \in \mathbb{R} .$$

Note that, in the case of a nonhomogeneous Poisson process, the local mean density and Bartlett density satisfy

$$\gamma^{<LS>}(\omega; u) = \frac{m_1^{<LS>}(u)}{2\pi} .$$

This is indeed given by (8) with a local fertility function to be identically zero, $p^{<LS>}(\cdot; u) \equiv 0$.

In Figures 1 and 2 we display the resulting estimators for the two assets ESSI.PA and TOTF.PA, respectively. Note that the scaling of the y -axis of the averaged local mean densities (top plots) is not the same. The transaction rate of ESSI.PA evolves around 0.1 transactions per second while that of the more liquid TOTF.PA around twice as much. The local Bartlett spectrum estimator $\hat{\gamma}_{b_2, b_1}(\omega; u)$ is computed over frequencies ω ranging between 0 and .1 Hz. This means that only cyclic behaviors with periods larger than 10 seconds are visible. As for the local mean density plots, note that the color scales of the averaged local Bartlett spectra are different for the two assets.

It is interesting to observe that, despite these differences of orders of magnitude, the shapes of the averaged local mean densities and that of the averaged local Bartlett spectra bear some similarities. Namely the averaged mean density is decreasing in the morning, although a sharp increase occurs around 11:00 a.m. and a drop during the lunch break. It then increases steadily during the afternoon with a sharper increase around 3:30 p.m., which corresponds to the opening time of the New York stock exchange market. The maximal averaged mean density is reached at the closing time. As for the averaged Bartlett spectrum, it is interesting to note that the shape of the spectrum along the frequencies varies significantly along the day. During the increases of mean density preceding and following lunch break, the spectrum concentrates at low frequencies, while the spectrum, although still favoring low frequencies, is more balanced during the increase following the opening of the NYSE market. Finally, it is interesting to observe that the Poisson-normalized Bartlett spectra take the highest values during the two one hour long periods surrounding the lunch break. It indicates that, in contrast to these two periods, the increase of the (nonnormalized) Bartlett spectrum toward the end of the day can be interpreted merely as a consequence of the increase of the local mean density rather than a departure from the Poisson behavior. Also observe that the Poisson-normalized Bartlett spectra are always larger than 1. Assuming a locally stationary Hawkes process for this data, this could be interpreted, according to Formula (8), as

$$|1 - P^{<LS>}(\omega; u)| < 1 ,$$

where $P^{<LS>}(\cdot; u)$ is the Fourier transform of the local fertility function $p^{<LS>}(\cdot; u)$.

A sensible conclusion of this analysis is that it advocates for more involved models than a simple non-homogeneous Poisson process for transaction data. In particular, locally

stationary Hawkes processes as assumed in this work are better adapted to such data sets, not only because the local Bartlett spectrum is not constant along the frequencies but also because its shape varies along the time, a feature that could not be obtained by using time varying baseline intensity with a fertility function constant over the time, as used in [Chen and Hall \[2013\]](#). This conclusion is of particular interest in relation with Examples 2.3 (iii) and (iv) described in [Dahlhaus and Tunyavetchakit \[2016\]](#) for modeling the volatility of high frequency financial data.

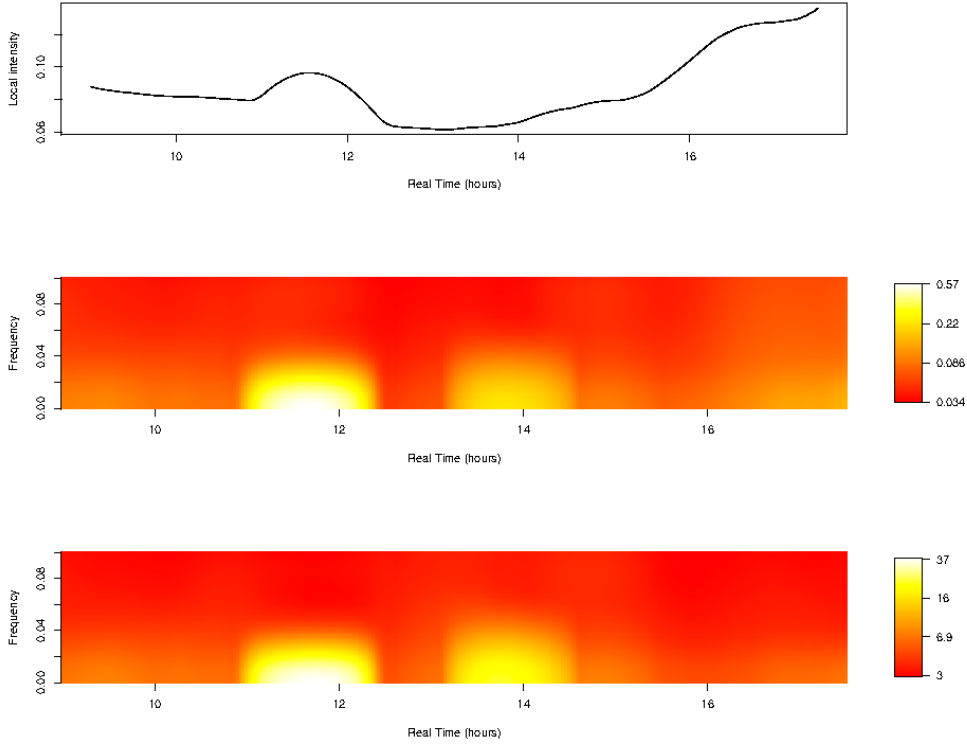


Figure 1: Averaged local mean density (top) and Bartlett spectrum, nonnormalized (middle) and Poisson-normalized (bottom), for ESSI.PA transaction data.

5 Deviation bounds for non-stationary Hawkes processes

We now derive new results required for treating the variance of the studied estimators. In contrast to Poisson processes, we can not rely on the independence of the process on disjoint sets. To the best of our knowledge, the most advanced results on deviation bounds of Hawkes processes are to be found in [Reynaud-Bouret and Roy \[2006\]](#) and only apply to stationary Hawkes processes with compactly supported fertility functions. Here we consider non-stationary Hawkes processes with exponentially decreasing local fertility functions. The generalization to non-stationary processes requires a specific approach rather than a mere adaption of [Reynaud-Bouret and Roy \[2006\]](#).

Although we are here motivated by the study of the variance of the local mean density and Bartlett estimators, we believe that the results contained in this section are of broader

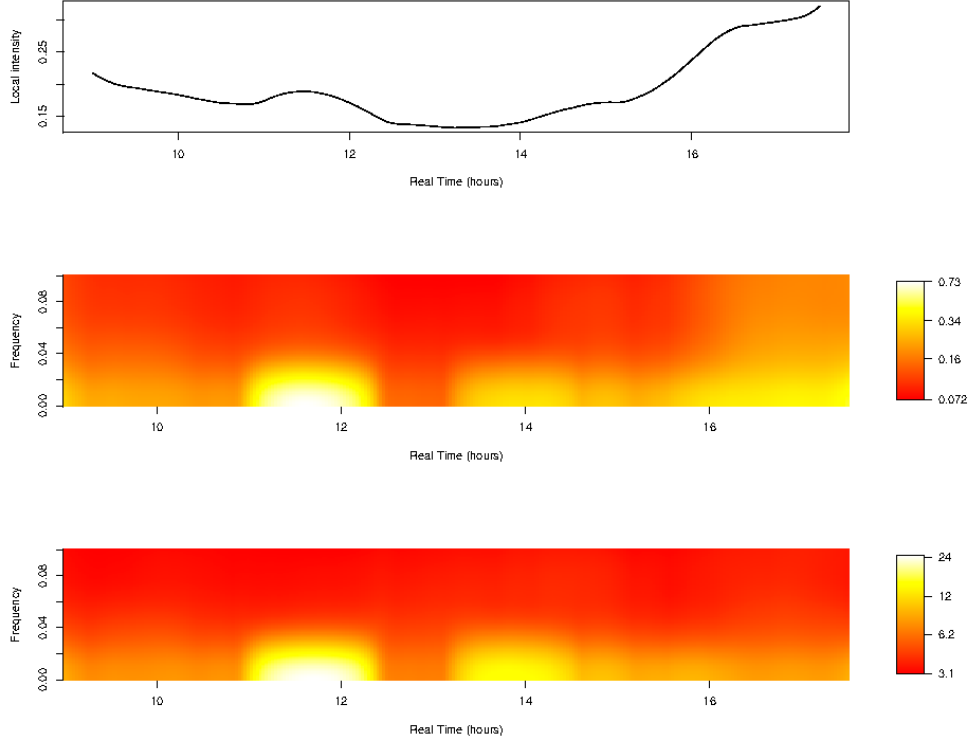


Figure 2: Averaged local mean density (top), Bartlett spectrum, nonnormalized (middle) and Poisson-normalized (bottom), for TOTF.PA transaction data.

interest, as they can serve more generally to understand the dependence structure of non-stationary Hawkes processes.

5.1 Setting

Recall that (5) is our minimal condition for defining the non-stationary Hawkes process N with immigrant intensity function $\lambda_c(\cdot)$ and varying fertility function $p(\cdot; \cdot)$. The cluster construction of a Hawkes process relies on conditioning on the realization of a so-called *center process*, N_c a Poisson point process (PPP) with intensity λ_c on \mathbb{R} , which describes spontaneous appearing of the immigrants. At each *center point* t^c of N_c , a point process $N(\cdot|t^c)$ is generated as a branching process whose conditional distribution given N_c only depends on t^c and is entirely determined by the time varying fertility function $p(\cdot; \cdot)$. This distribution is described through its conditional Laplace functional in [Roueff et al., 2016, section 6.1] under the additional assumption

$$\zeta_\infty < \infty \quad \text{with} \quad \zeta_q = \sup_{t \in \mathbb{R}} |p(\cdot; t)|_q \quad \text{for all } q \in [1, \infty] . \quad (44)$$

The following result directly follows from these derivations. A detailed proof is available in Appendix D for sake of completeness.

Proposition 6. *Suppose that (5) and (44) hold and set*

$$r_1 = \frac{-\log \zeta_1}{2(1 - \zeta_1^{1/2} + \zeta_\infty \zeta_1^{-1/2})} . \quad (45)$$

Then, for all $h : \mathbb{R} \rightarrow \mathbb{R}$ satisfying $|h|_1 \leq 1$ and $|h|_\infty \leq 1$,

$$\mathbb{E} \left[e^{(1-\zeta_1^{1/2})r_1|N(h)|} \right] \leq \exp \left(|\lambda_c|_\infty \zeta_1^{-1/2} r_1 \right) . \quad (46)$$

Consequently, for all $q > 0$, there exists a positive constant B_q only depending on $|\lambda_c|_\infty$, ζ_1 and ζ_∞ such that, for all $h : \mathbb{R} \rightarrow \mathbb{R}$ satisfying $|h|_1 \leq 1$ and $|h|_\infty \leq 1$,

$$\|N(h)\|_q \leq B_q . \quad (47)$$

The moment bound (47) will be useful but far from sufficient to bound the variance of our estimators efficiently. To see why, let us suppose temporarily that N is a homogeneous Poisson process with unit rate and consider $h = n^{-1} \mathbb{1}_{[0,n]} - 1$ for some positive integer n . Then the above bound with $p = 2$ gives $\text{Var}(N(h)) \leq B_2^2$ although we know that in this very special case we have $\text{Var}(N(h)) = n^{-1}$, hence the bound is not sharp at all as $n \rightarrow \infty$. In the following we provide new deviation bounds applying to non-stationary Hawkes processes which allows one to recover the expected order of magnitude for the variance. To this end we rely on stronger conditions than the ones used in Roueff et al. [2016].

Define moreover, using the exponentially weighted norm notation in (2), for all $d \geq 0$, and $q \in [1, \infty]$,

$$\zeta_q(d) = \sup_{t \in \mathbb{R}} |p(\cdot; t)|_{d,q} .$$

We strengthen the basic conditions (6) and (44) into the following one.

(NS-1) We have $|\lambda_c|_\infty < \infty$. Moreover, for all $t \in \mathbb{R}$, $p(\cdot; t)$ is supported on \mathbb{R}_+ and there exists $d > 0$ such that $\zeta_1(d) < 1$ and $\zeta_\infty(d) < \infty$.

The version of (NS-1) in the locally stationary case is (LS-1) in the sense that the locally stationary Hawkes process $(N_T)_{T \geq 1}$ satisfies (LS-1) if and only if, for all $T \geq 1$, the non-stationary Hawkes process N_T satisfied (NS-1). Therefore all the results below relying on (NS-1) apply to the locally stationary Hawkes processes satisfying (LS-1). We recall that this assumption means that the local fertility functions satisfy some uniform exponential decreasing.

5.2 New deviation bounds

The deviations bounds are based on the following exponential bound control on component point processes $N(\cdot|t^c)$ defined above.

Proposition 7. Suppose that (NS-1) holds for some $d > 0$. Then we have, for all $a \in (0, d)$, for all $t^c \in \mathbb{R}$,

$$\mathbb{E} \left[\int_{[t, \infty)} e^{a(s-t)} N(ds|t^c) \right] \leq C_a \quad \text{with} \quad C_a := 1 + \frac{\zeta_\infty(d)}{(d-a)(1-\zeta_1(d))} . \quad (48)$$

Proof. Let $g : \mathbb{R} \rightarrow \mathbb{R}_+$ and define, for all $h : \mathbb{R} \rightarrow \mathbb{R}_+$,

$$[\mathcal{E}_g(h)](s) = g(s) + \int h(u)p(u-s; u) du .$$

Following [Roueff et al., 2016, Proposition 7 and Eq. (34)], we have that, for all $t \in \mathbb{R}$,

$$\mathbb{E} [N(g|t^c)] = \lim_{n \rightarrow \infty} [\mathcal{E}_g^n(g)](t) , \quad (49)$$

where \mathcal{E}_g^n denotes the n -th self-composition of \mathcal{E}_g . In the following, we take some $a \in (0, d)$ and set

$$g_t(s) = e^{a(s-t)} \mathbb{1}_{[t, \infty)}(s) = g_0(s-t) .$$

Observe that, using that $p(\cdot; u)$ is supported on $[0, \infty)$ under (NS-1), for all $h : \mathbb{R} \rightarrow \mathbb{R}_+$,

$$\begin{aligned} [\mathcal{E}_{g_t}(h)](t) &\leq g_t(t) + \int_t^\infty h(u)p(u-t; u) \, du \\ &\leq 1 + \int_t^\infty h(u)e^{d(t-u)}e^{d(u-t)}p(u-t; u) \, du \\ &\leq 1 + \int_t^\infty h(u)e^{d(t-u)}\zeta_\infty(d) \, du \\ &= 1 + \zeta_\infty(d) |h(t+\cdot)|_{-d,1} . \end{aligned}$$

Applying this to $h = \mathcal{E}_{g_t}^{n-1}(g_t)$ we get, for all $n \geq 1$,

$$[\mathcal{E}_{g_t}^n(g_t)](t) \leq 1 + \zeta_\infty(d) |[\mathcal{E}_{g_t}^{n-1}(g_t)](t+\cdot)|_{-d,1} \quad (50)$$

Similarly, we also get that, for all $h : \mathbb{R} \rightarrow \mathbb{R}_+$,

$$|[\mathcal{E}_{g_t}(h)](t+\cdot)|_{-d,1} \leq |g_0|_{-d,1} + \zeta_1(d) |h(t+\cdot)|_{-d,1} .$$

Observing that $|g_0|_{-d,1} = (d-a)^{-1}$ and iterating the last inequality, we finally obtain that, for all $n \geq 1$,

$$|[\mathcal{E}_{g_t}^n(g_t)](t+\cdot)|_{-d,1} \leq \frac{1}{d-a} (1 + \zeta_1(d) + \dots + \zeta_1(d)^n) \leq \frac{1}{(d-a)(1-\zeta_1(d))} .$$

Inserting this bound in (50) and letting n go to ∞ as in (49) with $g = g_t$, we get the claimed result. \square

Define the past and future σ -fields at time t respectively as

$$\begin{aligned} \mathcal{F}_t &= \sigma(N_c(A), N(B|t^c) : A \in \mathbb{B}((-\infty, t]), B \in \mathbb{B}(\mathbb{R}), t^c \leq t) \\ &\supset \sigma(N(A) : A \in \mathbb{B}((-\infty, t])) \end{aligned}$$

and

$$\mathcal{G}_t = \sigma(N(A) : A \in \mathbb{B}((t, -\infty))) .$$

The following result provides a uniform exponential control on the time-dependence of N .

Proposition 8. *Suppose that (NS-1) holds for some $d > 0$ and that λ_c is bounded. Let $p \in [1, \infty)$, $t < u$ and Y be a centered L^1 \mathcal{G}_u -measurable random variable. Then, for all $a \in (0, d)$, for all $q \in (p, \infty]$, if Y is L^q ,*

$$\|\mathbb{E}[Y|\mathcal{F}_t]\|_p \leq \|Y\|_q \left(|\lambda_c|_\infty C_a a^{-1} e^{-a(u-t)} \right)^{1/p-1/q} ,$$

where C_a is defined in (48).

Proof. In the following, we denote by t_k^c the points of the Poisson process N_c , that is,

$$N_c = \sum_k \delta_{t_k^c} .$$

Define

$$\Delta_{t^c} = \inf \{ \delta > 0 : N([t^c + \delta, \infty)|t^c) = 0 \} .$$

In other words, Δ_{t^c} is the size of the cluster $N(\cdot|t^c)$, that is the distance between the most left point (which is t^c) and most right point. Since Δ_{t^c} is a point among the points of $N(\cdot|t^c)$, we have

$$e^{a\Delta_{t^c}} \leq \int_{[t, \infty)} e^{a(s-t)} N(ds|t^c) ,$$

and a direct consequence of Proposition 7 is that, for all $a \in (0, d)$,

$$\mathbb{E}[e^{a\Delta_{t^c}}] \leq C_a. \quad (51)$$

Now let us define the position of the last point generated by all clusters started before time t , namely,

$$\overline{\Delta}_t = \sup \{t_k^c + \Delta_{t_k^c} : t_k^c \leq t\}.$$

We observe that $\overline{\Delta}_t$ is \mathcal{F}_t -measurable. Moreover, if $t < u$ and Y is a centered $L^1 \mathcal{G}_u$ -measurable random variable, then we have $\mathbb{E}[Y|\mathcal{F}_t] = 0$ on $\{\overline{\Delta}_t < u\}$. The Hölder inequality then yields for $1 \leq p < q \leq \infty$

$$\|\mathbb{E}[Y|\mathcal{F}_t]\|_p = \left\| \mathbb{E}[Y|\mathcal{F}_t] \mathbb{1}_{\{\overline{\Delta}_t \geq u\}} \right\|_p \leq \|\mathbb{E}[Y|\mathcal{F}_t]\|_q (\mathbb{P}(\overline{\Delta}_t \geq u))^{1/p-1/q}.$$

Since $\|\mathbb{E}[Y|\mathcal{F}_t]\|_q \leq \|Y\|_q$, it only remains to prove that

$$\mathbb{P}(\overline{\Delta}_t \geq u) \leq C_0 e^{-\lambda_0(u-t)}. \quad (52)$$

Observe that $M = \sum_k \delta_{t_k^c, \Delta_{t_k^c}}$ is a marked Poisson point process such that, given N_c , the marks $\Delta_{t_k^c}$ are independent and for each k the conditional distribution of $\Delta_{t_k^c}$ only depends on t_k^c . Hence M is a Poisson point process with points valued in $\mathbb{R} \times \mathbb{R}_+$ and

$$\{\overline{\Delta}_t \geq u\} = \{M(\{(t^c, \delta) \in (-\infty, t] \times \mathbb{R}_+ : t_c + \delta \geq u\}) > 0\}.$$

We thus get that

$$\mathbb{P}(\overline{\Delta}_t \geq u) = 1 - \exp\left(-\int_{-\infty}^t \mathbb{P}(t_c + \Delta_{t^c} \geq u) \lambda_c(t^c) dt^c\right) \leq |\lambda_c|_\infty \int_{-\infty}^t \mathbb{P}(\Delta_{t^c} \geq u - t_c) dt^c,$$

where we used that $1 - e^{-x} \leq x$ for all $x \geq 0$. Using (51) and the exponential Markov inequality, it follows that

$$\mathbb{P}(\overline{\Delta}_t \geq u) \leq |\lambda_c|_\infty C_a \int_{-\infty}^t e^{a(t^c - u)} dt^c = |\lambda_c|_\infty C_a a^{-1} e^{a(t-u)}.$$

□

We can now derive a Burkholder-type inequality.

Proposition 9. *Suppose that (NS-1) holds for some $d > 0$. Let $p \in [2, \infty)$. Then there exists a positive constant B_p such that, for all bounded functions h with support included in $[j, j+n]$ for some $j \in \mathbb{Z}$ and $n \in \mathbb{N}$,*

$$\|N(h) - \mathbb{E}[N(h)]\|_p \leq A |h|_\infty \sqrt{n}.$$

where A is a positive constant only depending on d , $|\lambda_c|_\infty$, ζ_1 , ζ_∞ , $\zeta_\infty(d)$ and $\zeta_1(d)$, e.g., for any $a \in (0, d)$ and $q > p$,

$$A := (B_1 + B_p)(B_1 + B_q) (|\lambda_c|_\infty C_a a^{-1})^{1/p-1/q} \frac{e^{-a(1/p-1/q)}}{1 - e^{-a(1/p-1/q)}},$$

where B_p is defined in Proposition 6 and C_a in (48).

Proof. We can assume $|h|_\infty = 1$ without loss of generality. We write

$$h = \sum_{i=1}^n h_i \quad \text{with} \quad h_i = h \mathbb{1}_{[j+i-1, j+i]}.$$

Then $|h_i|_\infty \leq 1$ and $|h_i|_1 \leq 1$ for all i and, defining $X_i = N(h_i) - \mathbb{E}[N(h_i)]$, from Proposition 6, we have, for all $q \geq 1$,

$$\|X_i\|_q \leq B_q + B_1. \quad (53)$$

Then $N(h) - \mathbb{E}[N(h)] = \sum_{i=1}^n X_i$ and, applying [Dedecker et al., 2007, Proposition 5.4, Page 123], we have

$$\|N(h) - \mathbb{E}[N(h)]\|_p \leq \left(2p \sum_{i=1}^n b_{i,n} \right)^{1/2}, \quad (54)$$

where, denoting $\mathcal{M}_i = \mathcal{F}_{j+i}$,

$$b_{i,n} = \max_{1 \leq \ell \leq n} \left\| X_i \sum_{k=i}^{\ell} \mathbb{E}[X_k | \mathcal{M}_i] \right\|_{p/2}.$$

Observing that X_k is centered and \mathcal{G}_{j+k-1} -measurable, Proposition 8, gives that, for any $q > p$,

$$\|\mathbb{E}[X_k | \mathcal{M}_i]\|_p \leq \|X_k\|_q \left(|\lambda_c|_\infty C_a a^{-1} e^{-a(k-i+1)} \right)^{1/p-1/q}.$$

The Hölder inequality, the last two displays and (53) yield, for all $q > p$,

$$b_{i,n} \leq (B_1 + B_p)(B_1 + B_q) \left(|\lambda_c|_\infty C_a a^{-1} \right)^{1/p-1/q} \frac{e^{-a(1/p-1/q)}}{1 - e^{-a(1/p-1/q)}}.$$

Applying this in (54), we get the claimed bound. \square

Another consequence of Proposition 8 is the following useful covariance bound.

Corollary 10. *Suppose that (NS-1) holds for some $d > 0$ and that λ_c is bounded. Let h_1 and h_2 be two bounded integrable functions. Let γ satisfy one of the following assertions.*

- (i) *There exist $t \in \mathbb{R}$ such that $\text{Supp}(h_1) \subset (-\infty, t]$ and $\text{Supp}(h_2) \subset [t + \gamma, \infty)$.*
- (ii) *There exist $t \in \mathbb{R}$ such that $\text{Supp}(h_2) \subset (-\infty, t]$ and $\text{Supp}(h_1) \subset [t + \gamma, \infty)$.*
- (iii) $\gamma = 0$.

Then for all $q > 4$, there exists $C_q > 0$ and $\alpha_q > 0$ both only depending on q , $|\lambda_c|_\infty$, and a and C_a in Proposition 8 such that

$$\left| \text{Cov} \left(\left| \overline{N}(h_1) \right|^2, \left| \overline{N}(h_2) \right|^2 \right) \right| \leq C_q \left\| \overline{N}(h_1) \right\|_q^2 \left\| \overline{N}(h_2) \right\|_q^2 e^{-\alpha_q \gamma}, \quad (55)$$

where $\overline{N}(h) = N(h) - \mathbb{E}[N(h)]$.

Proof. In the case (iii), the bound (55) actually holds with $q = 4$ by the Cauchy-Schwarz inequality and thus also holds with $q > 4$ by Jensen's inequality.

We now consider the case (i) (the last one (ii) being obtained by inverting h_1 and h_2).

We have in this case, denoting $Y = \left| \overline{N}(h_2) \right|^2 - \mathbb{E} \left[\left| \overline{N}(h_2) \right|^2 \right]$,

$$\begin{aligned} \left| \text{Cov} \left(\left| \overline{N}(h_1) \right|^2, \left| \overline{N}(h_2) \right|^2 \right) \right| &= \mathbb{E} \left[\left| \overline{N}(h_1) \right|^2 (\mathbb{E}[Y | \mathcal{F}_t]) \right] \\ &\leq \left\| \overline{N}(h_1) \right\|_4^2 \left\| \mathbb{E}[Y | \mathcal{F}_t] \right\|_4^2. \end{aligned}$$

The Jensen Inequality and $q > 4$ give that $\left\| \overline{N}(h_1) \right\|_4 \leq \left\| \overline{N}(h_1) \right\|_q$ and the proof is concluded by using Proposition 8 to bound $\left\| \mathbb{E}[Y | \mathcal{F}_t] \right\|_4$. \square

Acknowledgements

We thank Adil Reghai of Natixis for providing us with the ESSI.PA and TOTF.PA transactions data sets. Rainer von Sachs gratefully acknowledges funding by contract “Projet d’Actions de Recherche Concertées” No. 12/17-045 of the “Communauté française de Belgique” and by IAP research network Grant P7/06 of the Belgian government (Belgian Science Policy). Both authors thank L. Sansonnet for her contributing discussions.

A Useful lemmas

In the following lemmas, we have gathered some simple bounds that we will repeatedly use in the sequel.

Lemma 11. *Let $a, \beta > 0$ and $p \in [1, \infty]$. For any function g and any $t \in \mathbb{R}$, we have*

$$|g(\cdot - t)|_{(\beta),p} \leq 2^{(\beta-1)+} \left(|g|_{(\beta),p} + |t|^\beta |g|_p \right), \quad (56)$$

$$|g(\cdot - t)|_{a,1} \leq e^{a|t|} |g|_{a,1}. \quad (57)$$

Let $q_{\omega_0, b_2} = b_2^{1/2} e^{i\omega_0 t} q(b_2 t)$ and $w_{Tb_1} = (Tb_1)^{-1} w(u/(Tb_1))$, where the kernels w and q satisfy $|w|_1 = |q|_2 = 1$. Then, we have, for all $b_1, b_2 \in (0, 1]$ and $T > 0$,

$$|q_{\omega_0, b_2}|_{(\beta),1} = b_2^{-1/2-\beta} |q|_{(\beta),1}, \quad (58)$$

$$|q_{\omega_0, b_2}|_1 = b_2^{-1/2} |q|_1, \quad (59)$$

$$|q_{\omega_0, b_2}|_2 = 1, \quad (60)$$

$$|q_{\omega_0, b_2}|_\infty = b_2^{1/2} |q|_\infty, \quad (61)$$

$$|q_{\omega_0, b_2}|_{(\beta),2} = b_2^{-\beta} |q|_{(\beta),2}, \quad (62)$$

$$|w_{Tb_1}|_1 = 1, \quad (63)$$

$$|w_{Tb_1}|_\infty = (b_1 T)^{-1} |w|_\infty, \quad (64)$$

$$|w_{Tb_1}|_{(\beta),1} = (b_1 T)^\beta |w|_{(\beta),1}, \quad (65)$$

$$|q_{\omega_0, b_2} * w_{Tb_1}|_1 \leq b_2^{-1/2} |q|_1, \quad (66)$$

$$|q_{\omega_0, b_2} * w_{Tb_1}|_2 \leq |w_{Tb_1}|_2 |q_{\omega_0, b_2}|_1 \leq b_2^{-1/2} (b_1 T)^{-1/2} |w|_2 |q|_1, \quad (67)$$

$$|q_{\omega_0, b_2} * w_{Tb_1}|_\infty \leq (b_1 T)^{-1} b_2^{-1/2} |w|_\infty |q|_1, \quad (68)$$

$$|q_{\omega_0, b_2} * w_{Tb_1}|_{(\beta),1} \leq 2^{(\beta-1)+} \left(b_2^{-1/2} (b_1 T)^\beta |q|_1 |w|_{(\beta),1} + b_2^{-1/2-\beta} |q|_{(\beta),1} \right). \quad (69)$$

If q and w have compact supports both included in $[-\tilde{a}, \tilde{a}]$ for some $\tilde{a} > 0$, we have

$$|w_{Tb_1}|_{a,1} = |w|_{aTb_1,1} \leq e^{a\tilde{a}Tb_1} |w|_1, \quad (70)$$

$$|q_{\omega_0, b_2}|_{a,1} = b_2^{-1/2} |q|_{a/b_2,1} \leq b_2^{-1/2} e^{a\tilde{a}b_2^{-1}} |q|_1, \quad (71)$$

$$|q_{\omega_0, b_2} * w_{Tb_1}|_{a,1} \leq |w_{Tb_1}|_{a,1} |q_{\omega_0, b_2}|_{a,1} \leq b_2^{-1/2} e^{a\tilde{a}(Tb_1+b_2^{-1})} |q|_1. \quad (72)$$

Proof. All these bounds are straightforward. We use the usual L^p bounds for convolution $|h \star g|_1 \leq |h|_1 |g|_1$ and $|h \star g|_2 \leq |h|_2 |g|_1$. When necessary, the weights are handled by using

$$|s|^\beta \leq 2^{(\beta-1)+} (|s-t|^\beta + |t|^\beta) \quad \text{and} \quad e^{a|s|} \leq e^{a|t|} e^{a|s-t|}. \quad (73)$$

□

The following lemma prepares the ground for deriving appropriate bounds used in the proof of Theorem 5.

Lemma 12. *Let $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}_+$ and $f : \mathbb{R} \rightarrow \mathbb{R}_+$ with $f \in L^1 \cap L^2$. Let moreover $f_\infty : \mathbb{R} \rightarrow \mathbb{R}_+$ satisfying, for all $s \in \mathbb{R}$,*

$$f_\infty(s) \leq f(s) + \int f_\infty(t) \varphi(s-t; t) dt. \quad (74)$$

Let us consider the following conditions depending on some $a \geq 0$ and $M \in (0, \infty]$.

$$(C-1) \quad \zeta_1(a) := \sup_t \int \varphi(u; t) e^{a|u|} du < 1.$$

$$(C-2) \quad \zeta_\infty(a) := \sup_{u,t} \varphi(u; t) e^{a|u|} < \infty.$$

$$(C-3) \quad \bar{\varphi}(u) := \sup_{|t| \leq M} \varphi(u; t) \text{ satisfies } \tilde{\zeta}_1 := \int \bar{\varphi}(u) du < 1.$$

If $M < \infty$, we define $\bar{f}_\infty : \mathbb{R} \rightarrow \mathbb{R}_+$ by $\bar{f}_\infty(s) := \int_{|t| > M} f_\infty(t) \varphi(s-t; t) dt$. Then the following assertions hold.

- (i) Condition (C-1) with $a \geq 0$ implies $|f_\infty|_{a,1} \lesssim |f|_{a,1}$.
- (ii) Condition (C-1) with $a > 0$ implies, for any $\beta \in (0, 1]$, $|f_\infty|_{(\beta),1} \lesssim |f|_{(\beta),1} + |f|_1$.
- (iii) Condition (C-3) with $M = \infty$ implies $|f_\infty|_2 \lesssim |f|_2$.
- (iv) Conditions (C-2) with $a > 0$ and (C-3) with $M = \infty$ imply $|f_\infty|_{(\beta),2} \lesssim |f|_{(\beta),2} + |f|_2$.
- (v) Conditions (C-1) with $a > 0$, (C-2) with $a = 0$ and (C-3) with $M < \infty$ imply $|f_\infty|_2 \lesssim |f|_2 + e^{-aM} |f|_{a,1}$.
- (vi) Conditions (C-1) and (C-2) with $a > 0$ and (C-3) with $M < \infty$ imply, for any $\beta \in (0, 1]$, $|f_\infty|_{(\beta),2} \lesssim |f|_{(\beta),2} + e^{-aM/2} |f|_{a,1} + |f|_2$.

Here " $\lesssim \dots$ " means " $\leq C \dots$ " with a positive constant C possibly depending on $a, \zeta_1(a), \zeta_\infty(a), \tilde{\zeta}_1$ or β only (thus neither depending on M nor on f).

Proof. Using (73) to deal with weighted L^p -norms, the bound (74) easily yields, for all $a \geq 0$ and $\beta \in (0, 1]$,

$$|f_\infty|_{a,1} \leq |f|_{a,1} + \zeta_1(a) |f_\infty|_{a,1}, \quad (75)$$

$$|f_\infty|_{(\beta),1} \leq |f|_{(\beta),1} + \left(\sup_t |\varphi(\cdot; t)|_{(\beta),1} \right) |f_\infty|_1 + \zeta_1(0) |f_\infty|_{(\beta),1}. \quad (76)$$

The bound (75) yields (i). Moreover, using, for any $\beta \in (0, 1]$, $C_{a,\beta} := \sup_{x \geq 0} x^\beta e^{-ax} \leq \sup_{x \geq 0} x e^{-ax} \leq (ea)^{-1}$, we get that

$$\sup_t |\varphi(\cdot; t)|_{(\beta),1} = \sup_t \int \varphi(r; t) |r|^\beta dr \leq \zeta_1(a) (ea)^{-1}.$$

Using this in (76) and (i) with $a = 0$, we get (ii).

The definition of $\bar{\varphi}$ in the case $M = \infty$ allows us to bound the second term of the bound in (74) by

$$\int f_\infty(t) \varphi(s-t; t) dt \leq f_\infty * \bar{\varphi}(s). \quad (77)$$

Using this, we get in turn that

$$|f_\infty|_2 \leq |f|_2 + |f_\infty * \bar{\varphi}|_2 \leq |f|_2 + |f_\infty|_2 |\bar{\varphi}|_1,$$

which under Condition (C-3) yields (iii).

Similarly to (76) and with the definition of $\bar{\varphi}$ in the case $M = \infty$, we have

$$|f_\infty|_{(\beta),2} \leq |f|_{(\beta),2} + |\bar{\varphi}|_{(\beta),1} |f_\infty|_2 + |\bar{\varphi}|_1 |f_\infty|_{(\beta),2} .$$

Note that in the case $M = \infty$, $|\bar{\varphi}|_{(\beta),1} \lesssim 1$ as a consequence of (C-2) with $a > 0$. Thus, using (iii) to bound $|f_\infty|_2$, under Condition (C-3), we get (iv).

Assertions (v) and (vi) are obtained similarly as (iii) and (iv) but with an additional step to deal with a finite M . Namely we have in this case

$$\int f_\infty(t) \varphi(s-t; t) dt \leq f_\infty * \bar{\varphi}(s) + \bar{f}_\infty(s) .$$

It follows that the same bounds as in (iii) and (iv) applies but with f replaced by $f + \bar{f}_\infty$. Hence to obtain the bounds in (v) and (vi), we only need to show that, under the corresponding conditions, we have

$$|\bar{f}_\infty|_2 \lesssim e^{-aM} |f|_{a,1} , \quad (78)$$

$$|\bar{f}_\infty|_{(\beta),2} \lesssim e^{-aM/2} |f|_{a,1} . \quad (79)$$

Observe that, by definition of \bar{f}_∞ , we have

$$\begin{aligned} |\bar{f}_\infty|_1 &= \int_{|t|>M} f_\infty(t) \left(\int \varphi(s-t; t) ds \right) dt \leq \zeta_1(0) \int_{|t|>M} f_\infty(t) dt , \\ |\bar{f}_\infty|_\infty &= \sup_s \int_{|t|>M} f_\infty(t) \varphi(s-t; t) dt \leq \zeta_\infty(0) \int_{|t|>M} f_\infty(t) dt . \end{aligned}$$

Using $|\bar{f}_\infty|_2 \leq (|\bar{f}_\infty|_\infty |\bar{f}_\infty|_1)^{1/2}$, $\int_{|t|>M} f_\infty(t) dt \leq e^{-aM} |f_\infty|_{a,1}$ and the bound in (i) to bound $|f_\infty|_{a,1}$ with $|f|_{a,1}$, we get (78).

Finally, we prove (79). First we note that, using (73), for $q = 1, \infty$, we have

$$|\bar{f}_\infty|_{(\beta),q} \leq \left(\sup_t |\varphi(\cdot; t)|_{(\beta),q} \right) \left(\int_{|t|>M} f_\infty(t) dt \right) + \left(\sup_t |\varphi(\cdot; t)|_q \right) \left(\int_{|t|>M} f_\infty(t) |t|^\beta dt \right) .$$

The bound (79) then follows similarly as (78) by using $|\bar{f}_\infty|_{(\beta),2} \leq (|\bar{f}_\infty|_{(\beta),\infty} |\bar{f}_\infty|_{(\beta),1})^{1/2}$ and (with the constant $C_{a/2,\beta}$ defined as above)

$$\int_{|t|>M} f_\infty(t) |t|^\beta dt \leq C_{a/2,\beta} e^{-aM/2} |f_\infty|_{a,1} \leq 2(ea)^{-1} e^{-aM/2} |f_\infty|_{a,1} .$$

This concludes the proof. \square

Lemma 13. *Let T, b_1, b_2 satisfy (21) and (22). Then for all $a_1, a_2 > 0$, we have*

$$\exp(a_2(Tb_1 + b_2^{-1}) - a_1T) \leq \max \left(1, \exp \left(2a_2 e^{1/a_1} \right) \right) .$$

In particular, under Convexion 2, we have $\exp(A_2(Tb_1 + b_2^{-1}) - A_1T) \lesssim 1$.

Proof. First note that, by (21), $b_2^{-1} \leq Tb_1$ and thus

$$\exp(a_2(Tb_1 + b_2^{-1}) - a_1T) \leq \exp(2a_2T(b_1 - a_1)) .$$

If $b_1 \leq a_1$ this upper bound is at most 1 and otherwise, using that $b_1 \leq 1$ and then (22), we have that $T(b_1 - a_1) \leq T \leq e^{1/b_1} \leq e^{1/a_1}$. \square

B Proof of Theorem 5

B.1 Preliminaries

In these preliminaries, we only require (5) and (44) to hold as they are sufficient to define a non-stationary Hawkes process. These two conditions are implied by (NS-1).

Let $g \in L^1 \cap L^\infty$, hence $g \in L^2$, too. Since N^c is a Poisson point process with intensity λ_c and the clusters $N(\cdot|t^c)$ can be seen as conditionally independent marks of this Poisson process, we have

$$\text{Var}(N(g)) = \int \text{Var}(N(g|t^c)) \lambda_c(t^c) dt^c + \int (\mathbb{E}[N(g|t^c)])^2 \lambda_c(t^c) dt^c \quad (80)$$

From [Corollary 8 and Proposition 11]Roueff et al. [2016] applied to the function $(s, z) \mapsto s g(z)$, we obtain that the applications $t \mapsto \mathbb{E}[N(g|t)]$ and $t \mapsto \text{Var}(N(g|t))$ are fixed points of two $L^1 \rightarrow L^1$ operators $h \mapsto \mathcal{E}(h)$ and $h \mapsto \tilde{\mathcal{E}}(h)$ defined as follows. The operator \mathcal{E} is defined by

$$\forall s \in \mathbb{R}, \quad \mathcal{E}(h)(s) = g(s) + \int h(t) p(t-s; t) dt. \quad (81)$$

And the operator $\tilde{\mathcal{E}}$ is defined similarly but with g replaced by the function

$$\tilde{g}(s) = \int (\mathbb{E}[N(g|t)])^2 p(t-s; t) dt.$$

By the first condition in (5), \mathcal{E} and $\tilde{\mathcal{E}}$ are strictly contracting and thus admit a unique fixed point. We denote this fixed point by g_∞ (resp. \tilde{g}_∞) in the following.

Hence, to summarize, the computation of $\text{Var}(N(g))$ boils down to the formula

$$\text{Var}(N(g)) = \int \tilde{g}_\infty(t) \lambda_c(t^c) dt^c + \int (g_\infty(t))^2 \lambda_c(t^c) dt^c, \quad (82)$$

where g_∞ and \tilde{g}_∞ are the unique fixed points of the $L^1 \rightarrow L^1$ operators \mathcal{E} and $\tilde{\mathcal{E}}$, with \mathcal{E} defined by (81) and $\tilde{\mathcal{E}}$ defined similarly but with $g(s)$ replaced by

$$\tilde{g}(s) = \int (g_\infty(t))^2 p(t-s; t) dt. \quad (83)$$

B.2 Decomposition of the approximation

The framework introduced in Section B.1 applies under the assumptions of Theorem 5 for computing both $\text{Var}(N_T(S^{-Tu}g))$ and $\text{Var}(N(g; u))$. For simplicity and without loss of generality we take $u = 0$ in the following and thus wish to approximate $\text{Var}(N_T(g))$ with $\text{Var}(N^{<S>}(g))$, with $N^{<S>}(g) := N(g; 0)$. Then $\text{Var}(N_T(g))$ and $\text{Var}(N(g; 0))$ satisfy Eq. (82) by adapting the definitions of \mathcal{E} , g_∞ , \tilde{g} and \tilde{g}_∞ to the corresponding $p(t-s; t)$ and λ_c . Namely, to compute $\text{Var}(N_T(g))$, we apply these equations and definitions with $p(t-s; t) = p^{<LS>}(t-s; t/T)$ and $\lambda_c(t^c) = \lambda_c^{<LS>}(t^c/T)$, while to compute $\text{Var}(N(g; 0))$, we apply them with $p(t-s; t) = p^{<S>}(t-s) := p^{<LS>}(t-s; 0)$ and $\lambda_c(t^c) = \lambda_c^{<S>} := \lambda_c^{<LS>}(0)$. To distinguish between these two cases, we use the corresponding notation $\mathcal{E}^{(T)}$, $g_\infty^{(T)}$, $\tilde{g}^{(T)}$ and $\tilde{g}_\infty^{(T)}$ in the first case and $\mathcal{E}^{<S>}$, $g_\infty^{<S>}$, $\tilde{g}^{<S>}$ and $\tilde{g}_\infty^{<S>}$ in the second case.

Applying (82) then yields the bound

$$|\text{Var}(N_T(g))| \leq |\lambda_c^{<LS>}|_\infty \left(\left| \tilde{g}_\infty^{(T)} \right|_1 + \left| g_\infty^{(T)} \right|_2^2 \right), \quad (84)$$

and the approximation bound

$$|\text{Var}(N_T(g)) - \text{Var}(N^{<S>}(g))| \leq (A) + (B) + (C) + (D)$$

with

$$\begin{aligned} (A) &= \int (g_\infty^{<S>}(t^c))^2 \left| \lambda_c^{<LS>}\left(\frac{t^c}{T}\right) - \lambda_c^{<S>} \right| dt^c, \\ (B) &= |\lambda_c^{<LS>}|_\infty \int \left| \left(g_\infty^{(T)}(t^c)\right)^2 - (g_\infty^{<S>}(t^c))^2 \right| dt^c, \\ (C) &= \int \tilde{g}_\infty^{<S>}(t^c) \left| \lambda_c^{<LS>}\left(\frac{t^c}{T}\right) - \lambda_c^{<S>} \right| dt^c, \\ (D) &= |\lambda_c^{<LS>}|_\infty \int \left| \tilde{g}_\infty^{(T)}(t^c) - \tilde{g}_\infty^{<S>}(t^c) \right| dt^c. \end{aligned}$$

Note that in (A) and (C), using (LS-2) we can bound $|\lambda_c^{<LS>}(t^c/T) - \lambda_c^{<S>}| = |\lambda_c^{<LS>}(t^c/T) - \lambda_c^{<LS>}(0)| \leq \xi_c^{(\beta)} T^{-\beta} |t^c|^\beta$. Hence, these four terms can be bounded using the previously introduced weighted norms, and we get

$$\begin{aligned} |\text{Var}(N_T(g)) - \text{Var}(N^{<S>}(g))| &\lesssim T^{-\beta} \left(|g_\infty^{<S>}|_{(\beta/2),2}^2 + |\tilde{g}_\infty^{<S>}|_{(\beta),1} \right) \\ &\quad + \left| \left(g_\infty^{(T)}\right)^2 - (g_\infty^{<S>})^2 \right|_1 + |\tilde{g}_\infty^{(T)} - \tilde{g}_\infty^{<S>}|_1. \end{aligned} \quad (85)$$

To bound these norms, we successively apply Lemma 12 in various settings.

B.3 Successive applications of Lemma 12

Norms involving $g_\infty^{<S>}$: We apply Lemma 12 with $f = |g|$, $\varphi(u; t) = \overline{\varphi}(u) = p^{<S>}(-u) = p^{<LS>}(-u; 0)$ and $f_\infty = |g_\infty^{<S>}|$. In this setting Eq. (74) is inherited from the fact that $g_\infty^{<S>}$ is a fixed point of $\mathcal{E}^{<S>}$. Conditions (C-1), (C-2) and (C-3) hold with $a = d > 0$ and $M = \infty$ as consequences of (LS-1).

Assertions (iii) and (iv) of Lemma 12 then respectively give

$$|g_\infty^{<S>}|_2 \lesssim |g|_2 \quad (86)$$

$$|g_\infty^{<S>}|_{(\beta/2),2} \lesssim |g|_{(\beta/2),2} + |g|_2. \quad (87)$$

Norms involving $g_\infty^{(T)}$: We apply Lemma 12 with $f = |g|$, $\varphi(s - t; t) = p^{<LS>}(t - s; t/T)$ and $f_\infty = |g_\infty^{(T)}|$. In this setting Eq. (74) is inherited from the fact that $g_\infty^{(T)}$ is a fixed point of $\mathcal{E}^{(T)}$. Conditions (C-1) and (C-2) hold with $a = d > 0$ as consequences of (LS-1). To check (C-3), we need to choose an appropriate $M < \infty$. From (LS-1) and (LS-3), we have $\overline{\varphi}(u) \leq p^{<LS>}(-u; 0) + \left(\frac{M}{T}\right)^\beta \xi^{(\beta)}(-u)$ and thus

$$\tilde{\zeta}_1 \leq \zeta_1^{<LS>}(0) + \left(\frac{M}{T}\right)^\beta \left| \xi^{(\beta)} \right|_1.$$

Since $\zeta_1^{<LS>}(0) \leq \zeta_1^{<LS>}(d) < 1$ and $|\xi^{(\beta)}|_1 < \infty$, we can define $\varepsilon > 0$ small enough, depending only on these two constants (hence $\varepsilon^{-1} \lesssim 1$), such that, if we set $M = \varepsilon T$ then we have $\tilde{\zeta}_1 \leq \zeta_1^{<LS>}(0)^{1/2} < 1$ and (C-3) follows.

Assertions (v) and (vi) of Lemma 12 then respectively give

$$\left| g_\infty^{(T)} \right|_2 \lesssim |g|_2 + e^{-d\varepsilon T} |g|_{d,1} \quad (88)$$

$$\left| g_\infty^{(T)} \right|_{(\beta),2} \lesssim |g|_{(\beta),2} + e^{-d\varepsilon T/2} |g|_{d,1} + |g|_2. \quad (89)$$

Norms involving $\tilde{g}_\infty^{<S>}$: Applying Lemma 12 (ii) with $f = |\tilde{g}^{<S>}|$ and $f_\infty = |\tilde{g}_\infty^{<S>}|$ and $\phi(u; t) = p^{<S>}(-u)$, we get that $|\tilde{g}_\infty^{<S>}|_{(\beta),1} \lesssim |\tilde{g}^{<S>}|_{(\beta),1} + |\tilde{g}^{<S>}|_1$. By definition of $\tilde{g}^{<S>}$ (adapted from (83) with $p(s-t; t) := p^{<S>}(s-t)$ and g_∞ replaced by $g_\infty^{<S>}$), we have $|\tilde{g}^{<S>}|_1 = |g_\infty^{<S>}|_2^2 |p^{<S>}|_1$ and, using (73), $|\tilde{g}^{<S>}|_{(\beta),1} = |g_\infty^{<S>}|_{(\beta/2),2}^2 |p^{<S>}|_1 + |g_\infty^{<S>}|_2^2 |p^{<S>}|_{(\beta),1}$. By (LS-1), we have $|p^{<S>}|_1, |p^{<S>}|_{(\beta),1} \lesssim 1$. Hence we finally get that

$$|\tilde{g}_\infty^{<S>}|_{(\beta),1} \lesssim |g_\infty^{<S>}|_2^2 + |g_\infty^{<S>}|_{(\beta/2),2}^2 . \quad (90)$$

Norms involving $\tilde{g}_\infty^{(T)}$: We proceed as in the previous case and get that $|\tilde{g}_\infty^{(T)}|_1 \lesssim |\tilde{g}^{(T)}|_1$ and $|\tilde{g}_\infty^{(T)}|_{(\beta),1} \lesssim |\tilde{g}^{(T)}|_{(\beta),1} + |\tilde{g}^{(T)}|_1$. Now $\tilde{g}^{(T)}$ is defined as in (83) with $p(s-t; t) := p^{<LS>}(s-t; t/T)$ and g_∞ replaced by $g_\infty^{(T)}$. We thus have $|\tilde{g}^{(T)}|_1 \leq \zeta_1^{<LS>}(0) |g_\infty^{(T)}|_2^2$ and, using (73),

$$|\tilde{g}^{(T)}|_{(\beta),1} \leq \zeta_1^{<LS>}(0) |g_\infty^{(T)}|_{(\beta/2),2}^2 + \left(\sup_r |p^{<LS>}(\cdot; r)|_{(\beta),1} \right) |g_\infty^{(T)}|_2^2 .$$

By (LS-1), we have $\sup_r |p^{<LS>}(\cdot; r)|_{(\beta),1} \lesssim 1$. Hence we finally get that

$$|\tilde{g}_\infty^{(T)}|_1 \lesssim |g_\infty^{(T)}|_2^2 , \quad (91)$$

$$|\tilde{g}_\infty^{(T)}|_{(\beta),1} \lesssim |g_\infty^{(T)}|_2^2 + |g_\infty^{(T)}|_{(\beta/2),2}^2 . \quad (92)$$

Norms involving $g_\infty^{(T)} - g_\infty^{<S>}$: Using that $g_\infty^{(T)}$ and $g_\infty^{<S>}$ are fixed points of $\mathcal{E}^{(T)}$ and $\mathcal{E}^{<S>}$, we find that $g_\infty := g_\infty^{<S>} - g_\infty^{(T)}$ satisfies

$$g_\infty(s) = \int g_\infty^{(T)}(t) (p^{<S>}(t-s) - p^{<LS>}(t-s; t/T)) dt + \int g_\infty(t) p^{<S>}(t-s) dt .$$

Hence taking absolute values $f_\infty := |g_\infty^{<S>} - g_\infty^{(T)}|$ satisfies (74) with

$$f(s) := \int |g_\infty^{(T)}(t)| |p^{<LS>}(t-s; t/T) - p^{<S>}(t-s)| dt .$$

and $\varphi(u; t) = p^{<S>}(-u)$. As previously Conditions (C-1), (C-2) and (C-3) hold with $a = d > 0$ and $M = \infty$ as consequences of (LS-1). Assertion (iii) of Lemma 12 then gives that $|g_\infty^{<S>} - g_\infty^{(T)}|_2 \lesssim |f|_2$ with f as in the previous display. By (LS-3), we further have that

$$|p^{<LS>}(t-s; t/T) - p^{<S>}(t-s)| \leq T^{-\beta} \xi^{(\beta)}(t-s) |t|^\beta , \quad (93)$$

and thus

$$|f|_2 = T^{-\beta} \left| \left(|g_\infty^{(T)}(\cdot)| \cdot |\cdot|^\beta \right) * \xi^{(\beta)} \right|_2 \leq T^{-\beta} |g_\infty^{(T)}|_{(\beta),2} |\xi^{(\beta)}|_1 . \quad (94)$$

Hence, with (89), we finally obtain that

$$|g_\infty^{<S>} - g_\infty^{(T)}|_2 \lesssim T^{-\beta} \left(|g|_{(\beta),2} + e^{-d\epsilon T/2} |g|_{d,1} + |g|_2 \right) . \quad (95)$$

Norms involving $\tilde{g}_\infty^{(T)} - \tilde{g}_\infty^{<S>}$: We apply the same line of reasoning as in the previous case. Using that $\tilde{g}_\infty^{(T)}$ and $\tilde{g}_\infty^{<S>}$ are fixed point of $\tilde{\mathcal{E}}^{(T)}$ and $\tilde{\mathcal{E}}^{<S>}$, we find that $f_\infty := |\tilde{g}_\infty^{<S>} - \tilde{g}_\infty^{(T)}|$ satisfies (74) with

$$f(s) := \left| \tilde{g}_\infty^{<S>}(s) - \tilde{g}_\infty^{(T)}(s) \right| + \int \left| \tilde{g}_\infty^{(T)}(t) \right| \left| p^{<S>}(t-s) - p^{<LS>}(t-s; t/T) \right| dt ,$$

and $\varphi(u; t) = p^{<S>}(-u)$. By definition of $\tilde{g}_\infty^{<S>}$ and $\tilde{g}_\infty^{(T)}$ (both adapted from (83)), we further have

$$\begin{aligned} \left| \tilde{g}_\infty^{<S>}(s) - \tilde{g}_\infty^{(T)}(s) \right| &\leq \int \left| g_\infty^{(T)}(t) \right|^2 \left| p^{<S>}(t-s) - p^{<LS>}(t-s; t/T) \right| dt \\ &\quad + \int \left| (g_\infty^{<S>}(t))^2 - (g_\infty^{(T)}(t))^2 \right| p^{<S>}(t-s) dt . \end{aligned}$$

Hence, using (93), we get that

$$\|f\|_1 \leq T^{-\beta} \left(\left| g_\infty^{(T)} \right|_{(\beta/2),2}^2 + \left| \tilde{g}_\infty^{(T)} \right|_1 \right) \left| \xi^{(\beta)} \right|_1 + \left| (g_\infty^{<S>})^2 - (g_\infty^{(T)})^2 \right|_1 \left| p^{<S>} \right|_1 .$$

Since $\left| p^{<S>} \right|_1, \left| \xi^{(\beta)} \right|_1 \lesssim 1$ under (LS-1) and (LS-3), Lemma 12 (i) with $a = 0$ thus yields

$$\left| \tilde{g}_\infty^{<S>} - \tilde{g}_\infty^{(T)} \right|_1 \lesssim T^{-\beta} \left(\left| g_\infty^{(T)} \right|_{(\beta/2),2}^2 + \left| \tilde{g}_\infty^{(T)} \right|_{(\beta),1} \right) + \left| (g_\infty^{<S>})^2 - (g_\infty^{(T)})^2 \right|_1 . \quad (96)$$

B.4 Conclusion of the proof

We can now gather the obtained bounds to conclude the proof of Theorem 5.

The bounds (84), (91) and (88) gives (31) (recall that $\varepsilon^{-1} \lesssim 1$).

Finally we prove (32). Using (85), and (96), we first obtain that

$$\begin{aligned} \left| \text{Var}(N_T(g)) - \text{Var}(N^{<S>}(g)) \right| &\lesssim T^{-\beta} (\text{I}) + \left| (g_\infty^{(T)})^2 - (g_\infty^{<S>})^2 \right|_1 , \\ \text{with } (\text{I}) &:= \left| g_\infty^{<S>} \right|_{(\beta/2),2}^2 + \left| \tilde{g}_\infty^{<S>} \right|_{(\beta),1} + \left| g_\infty^{(T)} \right|_{(\beta/2),2}^2 + \left| \tilde{g}_\infty^{(T)} \right|_{(\beta),1} . \end{aligned}$$

The bounds (90) and (92) and then (86), (87), (88), (89) (with $\beta/2$ instead of β) further give

$$\begin{aligned} (\text{I}) &\lesssim \left| g_\infty^{<S>} \right|_{(\beta/2),2}^2 + \left| g_\infty^{<S>} \right|_2^2 + \left| g_\infty^{(T)} \right|_{(\beta/2),2}^2 + \left| g_\infty^{(T)} \right|_2^2 \\ &\lesssim |g|_2^2 + |g|_{(\beta/2),2}^2 + e^{-A_1 T} |g|_{d,1}^2 . \end{aligned}$$

Using the Hölder inequality and then (95), (86) and (88), we get

$$\begin{aligned} \left| (g_\infty^{(T)})^2 - (g_\infty^{<S>})^2 \right|_1 &\leq \left| g_\infty^{(T)} - g_\infty^{<S>} \right|_2 \left(\left| g_\infty^{(T)} \right|_2 + \left| g_\infty^{<S>} \right|_2 \right) \\ &\lesssim T^{-\beta} \left(|g|_{(\beta),2} + e^{-A_1 T} |g|_{d,1} + |g|_2 \right) \left(|g|_2 + e^{-A_1 T} |g|_{d,1} \right) . \end{aligned}$$

Using that the products can be bounded by the sum of squares, the previous displays yield

$$\begin{aligned} \left| \text{Var}(N_T(g)) - \text{Var}(N^{<S>}(g)) \right| &\lesssim T^{-\beta} \left\{ |g|_{(\beta/2),2}^2 + |g|_2^2 + e^{-A_1 T} |g|_{d,1}^2 + |g|_{(\beta),2} \left(|g|_2 + e^{-A_1 T} |g|_{d,1} \right) \right\} . \end{aligned}$$

Now, by the Hölder inequality, we have $|g|_{(\beta/2),2}^2 \leq |g|_2 |g|_{(\beta),2}$, hence the first term inside the curly brackets can be removed by increasing the multiplicative constant by a factor 2 and we finally get (32), which concludes the proof of the theorem.

C Proof of main results

C.1 Proof of Theorem 1 (local mean density estimation)

For treating the bias, expressed as (34), we apply (29) in Theorem 4 with $g = w_{Tb_1}$. Using the norm estimates of equations (63) and (65), we immediately get

$$\mathbb{E}[\widehat{m}_{b_1}(u_0)] - m_1(u_0) \lesssim T^{-\beta} \left(1 + |w_{Tb_1}|_{(\beta),1}\right) \lesssim b_1^\beta + T^{-\beta}.$$

For treating the variance, expressed as (42), we use Proposition 9 with $h(\cdot) = w_{Tb_1}(\cdot - Tu_0)$ along with $|h|_\infty = |w_{Tb_1}|_\infty = (Tb_1)^{-1}$ by (64) and the obvious bound on the support of the kernel, $\text{Supp}(w_{Tb_1}) \lesssim Tb_1$. We immediately get (24), which concludes the proof.

C.2 Proof of Theorem 2 (Bias of spectral estimator)

The proof of this theorem requires to show two bounds, namely, the bound of the bias in time direction, (25), and the bound of the bias in frequency direction, (27). These two bounds are proved quite independently.

Proof of (25). The derivations of Section 3.3.2, namely (35), (36) and (37), show that we can decompose $\mathbb{E}[\widehat{\gamma}_{b_2,b_1}(u_0; \omega_0) - \gamma_{b_2}^{<LS>}(u_0; \omega_0)]$ as (I) + (II) - (III), where

$$\begin{aligned} \text{(I)} &= \int \left(\text{Var} \left(N_T(S^{-Tu_0} q_{\omega_0,b_2}(\cdot - t)) \right) - \gamma_{b_2}^{<LS>}(u_0; \omega_0) \right) w_{Tb_1}(t) dt \\ \text{(II)} &= \int \left| \mathbb{E}[\overline{N}_T(S^{-Tu_0} q_{\omega_0,b_2}(\cdot - t); u_0)] \right|^2 w_{Tb_1}(t) dt \\ \text{(III)} &= \underbrace{\text{Var} \left(N_T(S^{-Tu_0} q_{\omega_0,b_2} * w_{Tb_1}) \right)}_{\text{(IIIa)}} + \underbrace{\left| \mathbb{E}[\overline{N}_T(S^{-Tu_0} q_{\omega_0,b_2} * w_{Tb_1}; u_0)] \right|^2}_{\text{(IIIb)}} \end{aligned}$$

correspond to (39), (40) and (38), respectively. We will show now that

- (i) the term (I) is of order b_1^β ;
- (ii) the terms (II) and (IIIb) are of order $b_1^{2\beta} b_2^{-1}$;
- (iii) the term (IIIa) is of order $(Tb_1 b_2)^{-1}$;

which will conclude the proof of (25).

Term (I): By (32) in Theorem 5 with $g = q_{\omega_0,b_2}(\cdot - t)$, and recalling from equations (60), (62), (71), (56) and (57) that $|g|_2 = 1$, $|g|_{(\beta),2} \lesssim b_2^{-\beta} + |t|^\beta$ and

$$|g|_{d,1} \leq e^{d|t|} |q_{\omega_0,b_2}|_{d,1} \lesssim e^{d|t|} b_2^{-1/2} e^{A_2 b_2^{-1}} \lesssim e^{d|t|} e^{A_2 b_2^{-1}},$$

we have for all $t \in \mathbb{R}$,

$$\left| \text{Var} \left(N_T(S^{-Tu_0} q_{\omega_0,b_2}(\cdot - t)) \right) - \text{Var} \left(N(q_{\omega_0,b_2}; u_0) \right) \right| \lesssim T^{-\beta} \left(b_2^{-\beta} + |t|^\beta + e^{A_2|t|} e^{A_2 b_2^{-1} - A_1 T} \right)$$

where we have used that $1 \lesssim b_2^{-\beta} \lesssim e^{b_2^{-1}}$ and that $|t|^\beta \lesssim e^{|t|}$.

By (12), we can use this to bound the integrand in the definition of (I) and thus get, using (63), (65) and (70),

$$\text{(I)} \lesssim T^{-\beta} \left(b_2^{-\beta} + (Tb_1)^\beta + e^{-A_1 T + A_2(Tb_1 + b_2^{-1})} \right).$$

By (21) and applying Lemma 13, we have that the main term between the parentheses is the second one, hence we get (i).

Term (II): applying (29) in Theorem 4 with $g = q_{\omega_0, b_2}(\cdot - t)$ and using (58), (59) and (56), we get

$$\begin{aligned} |\mathbb{E}[\bar{N}_T(S^{-Tu_0} q_{\omega_0, b_2}(\cdot - t); u_0)]| &\lesssim T^{-\beta} \left(b_2^{-1/2}(1 + |t|^\beta) + b_2^{-1/2-\beta} \right) \\ &\lesssim T^{-\beta} \left(b_2^{-1/2}|t|^\beta + b_2^{-1/2-\beta} \right). \end{aligned} \quad (97)$$

(Since $b_2 \leq 1$) Taking the square and integrating this with respect to $w_{Tb_1}(t)dt$, and using (65), (63) and $T \geq 1$, we get (II) $\lesssim b_1^{2\beta} b_2^{-1}$ as claimed in (ii).

Term (IIIb): This term is treated similarly as (II) but using directly the bounds (66) and (69) inserted into (29) and taking the square. The same order is obtained and (ii) follows.

Term (IIIa): Applying the bound (31) of Theorem 5, we get that, setting here $g = q_{\omega_0, b_2} * w_{Tb_1}$,

$$\text{Var}(N_T(g)) \lesssim |g|_2^2 + e^{-aT} |g|_{d,1}^2.$$

Now, from (67) and (72), this bound reads

$$\text{Var}(N_T(g)) \lesssim b_2^{-1}(Tb_1)^{-1} + e^{-A_1 T + A_2(Tb_1 + b_2^{-1})}.$$

By Lemma 13 and since $Tb_1 b_2 \geq 1$, the first term dominates and we get (iii), which concludes the proof. \square

We now provide a proof for the second part of the theorem controlling the bias.

Proof of (27). This bound requires the usual control of the kernel-regularization of a smooth function as can be seen from (11). Namely, the function to consider is $\omega \mapsto \gamma^{<LS>}(\omega; u_0)$ and the kernel is $\omega \mapsto b_2^{-1} |Q((\omega - \omega_0)/b_2)|^2$ which integrates to 1 since $|Q|_2 = |q|_2 = 1$ by (K-2) and the Parseval theorem. Using the formula (8) to express $\omega \mapsto \gamma^{<LS>}(\omega; u_0)$ and the usual conditions on the kernel (26), it is thus sufficient to prove that, for all $\omega \in \mathbb{R}$,

$$\frac{m_1^{<LS>}(u_0)}{2\pi} \left| |1 - P^{<LS>}(\omega; u_0)|^{-2} - |1 - P^{<LS>}(\omega_0; u_0)|^{-2} + C(\omega - \omega_0) \right| \lesssim (\omega - \omega_0)^2,$$

where C is any constant (possibly depending on ω_0 and u_0 but not on depending on ω). As already seen, we have

$$m_1^{<LS>}(u_0) \leq |\lambda_c^{<LS>}|_\infty / (1 - \zeta_1^{<LS>}) \lesssim 1,$$

so that, we can consider the ratio $m_1^{<LS>}(u_0)/(2\pi)$ as a constant. For the remaining term involving the function $\omega \mapsto |1 - P^{<LS>}(\omega; u_0)|^{-2}$, we first observe that

$$\begin{aligned} P^{<LS>}(\omega; u_0) - P^{<LS>}(\omega_0; u_0) &= \int p^{<LS>}(t; u_0) (e^{-i\omega t} - e^{-i\omega_0 t}) dt \\ &= \tilde{p}(\omega_0; u_0) (\omega - \omega_0) + \int p^{<LS>}(t; u_0) e^{-i\omega_0 t} \left(e^{-i(\omega - \omega_0)t} - 1 - i(\omega - \omega_0)t \right) dt, \end{aligned}$$

with $\tilde{p}(\omega_0; u_0) := \int i t p^{<LS>}(t, u_0) e^{-i\omega_0 t} dt$. In the latter display, the first term is of the form $C(\omega - \omega_0)$ with $|C| \lesssim 1$ and the second term is of order $(\omega - \omega_0)^2$. This comes respectively from

$$|\tilde{p}(\omega_0; u_0)| \leq \int |t| p^{<LS>}(t; u_0) dt \leq \zeta_1^{<LS>} + \zeta_{(2)}^{<LS>}$$

and

$$\left| \int p^{<\text{LS}>}(t, u_0) e^{-i\omega_0 t} \left(e^{-i(\omega - \omega_0)t} - 1 - i(\omega - \omega_0)t \right) dt \right| \leq \zeta_{(2)}^{<\text{LS}>} (\omega - \omega_0)^2.$$

To conclude the proof we argue that the form $C(\omega - \omega_0) + R(\omega)$ with $C \lesssim 1$ and $R(\omega) \lesssim (\omega - \omega_0)^2$ satisfied by $P^{<\text{LS}>}(\omega; u_0) - P^{<\text{LS}>}(\omega_0; u_0)$ is inherited by $|1 - P^{<\text{LS}>}(\omega; u_0)|^{-2} - |1 - P^{<\text{LS}>}(\omega_0; u_0)|^{-2}$. This follows from

$$|1 - P^{<\text{LS}>}(\omega; u_0)|^{-1} \leq (1 - \zeta_1^{<\text{LS}>})^{-1} \in (1, \infty)$$

(using (6)) and from the identity valid for all complex numbers z, z' inside the unit disk

$$\frac{1}{|1 - z|^2} - \frac{1}{|1 - z'|^2} = \frac{|1 - z|^2 - |1 - z'|^2}{|1 - z'|^4} + \frac{(|1 - z|^2 - |1 - z'|^2)^2}{|1 - z'|^4 |1 - z|^2}$$

The numerator of the first term of this sum can be expressed as a sum of terms depending on the difference $|z - z'|$, and this applies to the numerator of the second term, too (as it is the square of the first numerator):

$$|1 - z|^2 - |1 - z'|^2 = 2\Re(z - z')\Re(1 - z') + 2\Im(z - z')\Im(1 - z') + |z - z'|^2.$$

□

C.3 Additional lemmas

As explained in Section 3.3.3, the treatment of the variance is done via the introduction of the “truly” centered process $\overline{\overline{N}}_T$. Here we provide three lemmas, two concerned with useful bounds for this centered process and the third one which controls the quality of the approximation of the estimator based on the centered process.

Lemma 14. *Let $p \geq 1$. Under the conditions of Theorem 3, we have, for all $u_0 \in \mathbb{R}$, $\omega_0 \in \mathbb{R}$, and b_1, b_2, T as in (21),*

$$\left\| \overline{\overline{N}}_T(S^{-T u_0} q_{\omega_0, b_2} * w_{T b_1}) \right\|_p \lesssim (T b_1 b_2)^{-1/2}, \quad (98)$$

where $\overline{\overline{N}}_T$ is defined in (41). Let moreover $h : \mathbb{R} \rightarrow \mathbb{C}$ be such that, for all $t \in \mathbb{R}$, we have

$$|h(t)| \leq (a_T + b_T |t|^\beta) w_{T b_1}(t),$$

for two positive constants a_T and b_T (possibly depending on T, b_1 and b_2). Then we have, for all $u_0 \in \mathbb{R}$, $\omega_0 \in \mathbb{R}$, and b_1, b_2, T as in (21),

$$\left\| \overline{\overline{N}}_T(S^{-T u_0} q_{\omega_0, b_2} * h) \right\|_p \lesssim (a_T + b_T (b_1 T)^\beta) (T b_1 b_2)^{-1/2}. \quad (99)$$

Proof. We apply Proposition 9, and get

$$\left\| \overline{\overline{N}}_T(S^{-T u_0} q_{\omega_0, b_2} * w_{T b_1}) \right\|_p \leq A |q_{\omega_0, b_2} * w_{T b_1}|_\infty \sqrt{n},$$

for some generic constant A and with a positive integer upper bound n on the length of the support of $S^{-T u_0} q_{\omega_0, b_2} * w_{T b_1}$, denoted by $\text{Supp}(S^{-T u_0} q_{\omega_0, b_2} * w_{T b_1})$. Observing that $\text{Supp}(S^{-T u_0} q_{\omega_0, b_2} * w_{T b_1}) \subset \text{Supp}(S^{-T u_0} q_{\omega_0, b_2}) + \text{Supp}(w_{T b_1})$, that the length of $\text{Supp}(q_{\omega_0, b_2})$ is of order b_2^{-1} and that the length of $\text{Supp}(w_{T b_1})$ is of order $T b_1$, we have $n \lesssim b_2^{-1} + T b_1 \lesssim T b_1$. We thus obtain the bound (98) with (68).

The bound (99) is obtained similarly but this time we rely on the bound

$$|q_{\omega_0, b_2} * h|_\infty \leq a_T |q_{\omega_0, b_2} * w_{T b_1}|_\infty + b_T |q_{\omega_0, b_2}|_1 |w_{T b_1}|_\infty n^\beta.$$

(Recall that n is length of the support of $w_{T b_1}$.) With (68), $n \lesssim T b_1$, (61) and (65) we get $|q_{\omega_0, b_2} * h|_\infty \lesssim b_2^{-1/2} (b_1 T)^{-1} (a_T + b_T (b_1 T)^\beta)$, which yields (99). □

Lemma 15. *Under the conditions of Theorem 3, we have, for all $u_0 \in \mathbb{R}$, $\omega_0 \in \mathbb{R}$, and b_1, b_2, T as in (21),*

$$\text{Var} \left(\int |\overline{N}_T(S^{-Tu_0} q_{\omega_0, b_2}(\cdot - t))|^2 w_{Tb_1}(t) dt \right) \lesssim (Tb_1 b_2)^{-1}. \quad (100)$$

Proof. We can write the left-hand side of (100) as

$$\iint \text{Cov} \left(|\overline{N}_T(f(\cdot - t))|^2, |\overline{N}_T(f(\cdot - t'))|^2 \right) w_{Tb_1}(t) w_{Tb_1}(t') dt dt'.$$

Let ℓ denote the length of the support of q_{ω_0, b_2} , which clearly satisfies for $b_2 \in (0, 1]$,

$$\ell \lesssim b_2^{-1}. \quad (101)$$

Then with $h_1 = q_{\omega_0, b_2}(\cdot - t)$ and $h_2 = q_{\omega_0, b_2}(\cdot - t')$, setting $\gamma := (|t - t'| - \ell)_+$, we have one of the assertions (i), (ii) or (iii) which is satisfied. Hence Corollary 10 gives that, for some $q > 4$,

$$\left| \text{Cov} \left(\left| \overline{N}_T(h_1) \right|^2, \left| \overline{N}_T(h_2) \right|^2 \right) \right| \leq C_q \left\| \overline{N}_T(h_1) \right\|_q^2 \left\| \overline{N}_T(h_2) \right\|_q^2 e^{-\alpha_q(|t-t'|-\ell)_+}. \quad (102)$$

Further we apply Proposition 9 with (61) and (101) and get, for $i = 1, 2$, $\left\| \overline{N}_T(h_i) \right\|_q^2 \lesssim (|q_{\omega_0, b_2}|_\infty)^2 \ell \lesssim 1$. Hence we finally get

$$\begin{aligned} \text{Var} \left(\int |\overline{N}_T(S^{-Tu_0} q_{\omega_0, b_2}(\cdot - t))|^2 w_{Tb_1}(t) dt \right) &\lesssim \iint e^{-\alpha_q(|t-t'|-\ell)_+} w_{Tb_1}(t) w_{Tb_1}(t') dt dt' \\ &\leq |w_{Tb_1} * w_{Tb_1}|_\infty \int e^{-\alpha_q(|u|-\ell)_+} du. \end{aligned}$$

Now, by (101) we have $\int e^{-\alpha_q(|u|-\ell)_+} du \lesssim b_2^{-1}$ and by (63) and (64), we have $|w_{Tb_1} * w_{Tb_1}|_\infty \lesssim (b_1 T)^{-1}$. Hence we get (100) and the proof is concluded. \square

Lemma 16. *Under the conditions of Theorem 3, we have, for all $u_0 \in \mathbb{R}$, $\omega_0 \in \mathbb{R}$, and b_1, b_2, T as in (21),*

$$\|\widehat{\gamma}_{b_2, b_1}(u_0; \omega_0) - \widetilde{\gamma}_{b_2, b_1}(u_0; \omega_0)\|_2 \lesssim b_1^{2\beta} b_2^{-1} + (Tb_1 b_2)^{-1}, \quad (103)$$

where $\widehat{\gamma}_{b_2, b_1}(u_0; \omega_0)$ and $\widetilde{\gamma}_{b_2, b_1}(u_0; \omega_0)$ are respectively defined by (13) and (43).

Proof. By definitions (33) and (41), we have, for any integrable test function f , $\overline{N}_T(f) = \overline{N}_T(f; u_0) - \mathbb{E}[\overline{N}_T(f; u_0)]$. Thus, (35) and (37) yield the following expression for $\widehat{\gamma}_{b_2, b_1}(u_0; \omega_0)$

$$\int \left| \overline{N}_T(f(\cdot - t)) + \mathbb{E}[\overline{N}_T(f(\cdot - t); u_0)] \right|^2 w_{Tb_1}(t) dt - \left| \overline{N}_T(f * w_{Tb_1}) + \mathbb{E}[\overline{N}_T(f * w_{Tb_1}; u_0)] \right|^2,$$

where we used the test function $f = S^{-Tu_0} q_{\omega_0, b_2}$. Developing the first square modulus and using for the second that $|a + b|^2 \leq 2|a|^2 + 2|b|^2$, and since $\int w_{Tb_1} = 1$, we obtain

$$|\widehat{\gamma}_{b_2, b_1}(u_0; \omega_0) - \widetilde{\gamma}_{b_2, b_1}(u_0; \omega_0) - 2\Re(B_T)| \leq A_T + 2C_T + 2D_T, \quad (104)$$

where we set, denoting by f^* the complex conjugate of f ,

$$\begin{aligned} A_T &:= \int |\mathbb{E}[\overline{N}_T(f(\cdot - t); u_0)]|^2 w_{Tb_1}(t) dt, \\ B_T &:= \int \overline{N}_T(f(\cdot - t)) \mathbb{E}[\overline{N}_T(f^*(\cdot - t); u_0)] w_{Tb_1}(t) dt, \\ C_T &:= \left| \overline{N}_T(f * w_{Tb_1}) \right|^2 \quad \text{and} \quad D_T := \left| \mathbb{E}[\overline{N}_T(f * w_{Tb_1}; u_0)] \right|^2. \end{aligned}$$

Note that A_T and D_T have been treated in the proof of Theorem 2 as the (deterministic) terms (II) and (IIIb). The assumptions in Theorem 2 are weaker than that of this lemma. Hence we can directly use (ii) of the proof of Theorem 2 and obtain that

$$A_T, D_T \lesssim b_1^{2\beta} b_2^{-1}.$$

The bound (98) in Lemma 14 immediately gives

$$\|C_T\|_2 = \left\| \overline{N}_T(S^{-T u_0} q_{\omega_0, b_2} * w_{T b_1}) \right\|_4^2 \lesssim (T b_1 b_2)^{-1}.$$

and we are left with treating B_T . Note that we can rewrite B_T as $B_T = \overline{N}_T(f * h)$, where h is the (deterministic) function $t \mapsto \mathbb{E}[\overline{N}_T(f^*(\cdot - t); u_0)] w_{T b_1}(t)$. Now, by (97), we have, for all $t \in \mathbb{R}$,

$$|h(t)| \lesssim b_2^{-1/2} T^{-\beta} (b_2^{-\beta} + |t|^\beta) w_{T b_1}(t).$$

Applying (99) with $a_T = b_2^{-1/2} (b_2 T)^{-\beta}$ and $b_T = b_2^{-1/2} T^{-\beta}$ in Lemma 14, it follows that

$$\|B_T\|_2 \lesssim b_2^{-1/2} T^{-\beta} (b_2^{-\beta} + (b_1 T)^\beta) (T b_1 b_2)^{-1/2} \lesssim b_2^{-1/2} b_1^\beta (T b_1 b_2)^{-1/2},$$

for $T b_1 b_2 \geq 1$.

Inserting the previous bounds on A_T , B_T , C_T and D_T in (104) yields

$$\|\widehat{\gamma}_{b_2, b_1}(u_0; \omega_0) - \widetilde{\gamma}_{b_2, b_1}(u_0; \omega_0)\|_2 \lesssim b_2^{-1/2} b_1^\beta (T b_1 b_2)^{-1/2} + b_1^{2\beta} b_2^{-1} + (T b_1 b_2)^{-1}.$$

Using that $2b_2^{-1/2} b_1^\beta (T b_1 b_2)^{-1/2} \leq b_1^{2\beta} b_2^{-1} + (T b_1 b_2)^{-1}$, we get (103). \square

C.4 Proof of Theorem 3 (variance of spectral estimator)

Lemmas 15 and 16, together with the definition of $\widetilde{\gamma}_{b_2, b_1}(u_0; \omega_0)$ in (43), directly give that

$$\text{Var}(\widehat{\gamma}_{b_2, b_1}(u_0; \omega_0)) \lesssim \frac{1}{T b_1 b_2} + \left(b_1^{2\beta} b_2^{-1} + (T b_1 b_2)^{-1} \right)^2.$$

Since $T b_1 b_2 \geq 1$ in (21), the second term within the squared parentheses is at most of the same order as the term outside the squared parentheses and can thus be discarded. Hence we obtain Theorem 3.

D Proof of Proposition 6

This proof uses the notation and definitions of [Roueff et al., 2016, Section 2.1]. Consider any

$$r_\infty \in (0, -\log \zeta_1) \quad \text{and} \quad r_1 \in (0, r_\infty e^{-r_\infty} \zeta_\infty^{-1}). \quad (105)$$

and set

$$R_1 := r_1 (1 - \zeta_1 e^{r_\infty}) \in (0, r_1), \quad (106)$$

$$R_\infty := r_\infty - e^{r_\infty} \zeta_\infty r_1 \in (0, r_\infty). \quad (107)$$

Let $K \subset U$ be a compact set and $g \in B_{\mathcal{O}}(R_1; K, 1) \cap B_{\mathcal{O}}(R_\infty; K, \infty)$.

Corollary 12 and Eq. (33) in Roueff et al. [2016] give that if $g \in B_{\mathcal{O}}(R_1; K, 1) \cap B_{\mathcal{O}}(R_\infty; K, \infty)$, with R_1, R_∞ defined by (106) and (107), we have, for all $z \in K$,

$$\mathcal{L}(g(\cdot, z)) := \mathbb{E} \left[e^{N(g(\cdot, z))} \right] = \exp \int (\exp(\Phi_g^\infty(t^c, z)) - 1) \lambda_c(t^c) dt^c,$$

where Φ_g^∞ is defined in [Roueff et al., 2016, Definition 3] as an element of $B_{\bar{\mathcal{O}}}(r_1; K, 1) \cap B_{\bar{\mathcal{O}}}(r_\infty; K, \infty)$.

Let now $h : \mathbb{R} \rightarrow \mathbb{R}_+$ be a bounded and integrable function and set $g(t, z) = z h(t)$. Let $U = \mathbb{R}$ and $K = [-r, r]$ for some $r > 0$. The previous display and the bound $|e^a - 1| \leq |a|e^{|a|}$ give that

$$\sup_{|z| \leq r} \left| \mathbb{E} \left[e^{z N(h)} \right] \right| \leq \exp \left(|\lambda_c|_\infty e^{|\Phi_g^\infty|_{\bar{\mathcal{O}}, K, \infty}} |\Phi_g^\infty|_{\bar{\mathcal{O}}, K, 1} \right) \leq \exp(|\lambda_c|_\infty e^{r_\infty} r_1) .$$

Here r_1 and r_∞ should be taken as small as possible provided that (105) holds and $g \in B_{\bar{\mathcal{O}}}(R_1; K, 1) \cap B_{\bar{\mathcal{O}}}(R_\infty; K, \infty)$, with R_1, R_∞ defined by (106) and (107). The specific choice of g and K here gives $g \in B_{\bar{\mathcal{O}}}(R_1; K, 1) \cap B_{\bar{\mathcal{O}}}(R_\infty; K, \infty)$ if

$$r |h|_1 \leq R_1 \quad \text{and} \quad r |h|_\infty \leq R_\infty .$$

So we conclude that

$$\sup_{|z| \leq r} \left| \mathbb{E} \left[e^{z N(h)} \right] \right| \leq \exp(|\lambda_c|_\infty e^{r_\infty} r_1) ,$$

for any r_1 and r_∞ satisfying (105) and r satisfying

$$r < \min \left(\frac{r_1 (1 - \zeta_1 e^{r_\infty})}{|h|_1}, \frac{r_\infty - e^{r_\infty} \zeta_\infty r_1}{|h|_\infty} \right) .$$

Let us set $r_\infty = (-\log \zeta_1)/2$ so that (105) reduces to

$$0 < r_1 < (-\log \zeta_1) \zeta_1^{1/2} \zeta_\infty^{-1} / 2 .$$

In the particular case where $|h|_1 \leq 1$ and $|h|_\infty \leq 1$, the condition on r then reads as

$$r < \min(r_1(1 - \zeta_1^{1/2}), (-\log \zeta_1)/2 - r_1 \zeta_\infty \zeta_1^{-1/2}) = r_1(1 - \zeta_1^{1/2}) ,$$

where the last equality holds for the choice of r_1 given by (45). We thus get, for all $r < r_1(1 - \zeta_1^{1/2})$,

$$\mathbb{E} \left[e^{r N(h)} \right] \leq \exp \left(|\lambda_c|_\infty \zeta_1^{-1/2} r_1 \right) .$$

Letting r tend to $r_1(1 - \zeta_1^{1/2})$, we also get the result for $r = r_1(1 - \zeta_1^{1/2})$ which corresponds to (46) for a non-negative h . To conclude, if h is signed we use that $|N(h)| \leq N(|h|)$ and apply the previous bound to $|h|$.

References

- O. Aalen. *Statistical inference for a family of counting processes*. PhD thesis, University of California, Berkeley, 1975.
- P. K. Andersen, Ø. Borgan, N. L. Hjort, E. Arjas, J. Stene, and O. Aalen. Counting process models for life history data: A review [with discussion and reply]. *Scandinavian Journal of Statistics*, 12(2):97, 1985. ISSN 03036898.
- E. Bacry, S. Delattre, M. Hoffmann, and J.F. Muzy. Modelling microstructure noise with mutually exciting point processes. *Quant. Finance*, 13(1):65–77, 2013.
- S. Birr, S. Volgushev, T. Kley, H. Dette, and M. Hallin. Quantile spectral analysis for locally stationary time series. to appear in the Journal of the Royal Statistical Society: Series B, 2014.

- F. Chen and P. Hall. Inference for a non-stationary self-exciting point process with an application in ultra-high frequency financial data modeling. *J. Appl. Probab.*, 50:1006–1024, 2013.
- R. Dahlhaus. Asymptotic statistical inference for nonstationary processes with evolutionary spectra. In *Athens Conference on Applied Probability and Time Series Analysis, Vol. II (1995)*, volume 115 of *Lecture Notes in Statist.*, pages 145–159. Springer, New York, 1996a.
- R. Dahlhaus. On the Kullback-Leibler information divergence of locally stationary processes. *Stochastic Process. Appl.*, 62:139–168, 1996b.
- R. Dahlhaus. A likelihood approximation for locally stationary processes. *Ann. Statist.*, 28(6):1762–1794, 2000. ISSN 0090-5364.
- R. Dahlhaus and S. Tunyavetchakit. Volatility decomposition and estimation in time-changed price models. Technical report, Heidelberg, 2016.
- D.J. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes - Volume I: Elementary Theory and Methods*. Probability and its Applications (New York). Springer, second edition, 2003.
- J. Dedecker, P. Doukhan, G. Lang, J. R. León, S. Louhichi, and C. Prieur. *Weak dependence: with examples and applications*, volume 190 of *Lecture Notes in Statistics*. Springer, New York, 2007. ISBN 978-0-387-69951-6.
- N.R. Hansen, P. Reynaud-Bouret, and V. Rivoirard. Lasso and probabilistic inequalities for multivariate point processes. *Bernoulli*, 21(1):83–143, 2015.
- H. Ramlau-Hansen. Smoothing counting process intensities by means of kernel functions. *Ann. Statist.*, 11(2):453–466, 1983. ISSN 0090-5364. doi: 10.1214/aos/1176346152.
- P. Reynaud-Bouret and E. Roy. Some non asymptotic tail estimates for Hawkes processes. *Bull. Belg. Math. Soc. Simon Stevin*, 13:883–896, 2006.
- F. Roueff, R. von Sachs, and L. Sansonnet. Locally stationary Hawkes processes. *Stochastic Processes and their Applications*, 126(6):1710 – 1743, 2016.
- B. Zheng, F. Roueff, and F. Abergel. Modelling bid and ask prices using constrained Hawkes processes: Ergodicity and scaling limit. *SIAM J. Finan. Math.*, 5(1):99–136, February 2014. doi: 10.1137/130912980.
- Z. Zhou and W.B. Wu. Local linear quantile estimation for non-stationary time series. *Ann. Statist.*, 37:2696–2729, 2009.